

KCI 문헌 유사도 검사 결과 확인서

* 유의사항

KCI 문헌 유사도 검사에서 나타나는 유사도 수치는 단순한 자동검사 결과이므로,
문헌 간 유사여부 판단을 위해서는 반드시 해당 분야 전문가의 직접 검사가 필요함을 알려드립니다.

확 인

유사율		1%	
발급번호	00013837576	발급일자	2026.01.26 19:51
검사일자	2026.01.26 19:50		
검사명	Purchase Forecasting AI		
검사문서	Article_20260126_PurchaseForecastingAI_JLKK.docx		
비고			
검사설정	유사율 기준 [5이절], 인용문장 [포함], 출처표시문장 [포함], 목차/참고문헌 [제외]		
비교범위	[KCI 논문] [학술대회 논문]		

유사 분석 정보(상세)					
문서유사율	전체문장	동일문장	유사의심문장	인용포함문장	출처표시문장
1%	510	0	11	6	34

비교 문서 정보				
번호	유사율	출처정보		비고
1	1%	[KCI 논문] KCI 논문 - 제목 : 설명가능한 AI 기반 여자 기계체조 마루운동 심판 점수의 패턴 분석 및 편향 시각화 - 저자 : 윤지운(한국체육대학교) - 발행년 : 2025.10		
2	1%	[KCI 논문] KCI 논문 - 제목 : 청소년기 디지털 활용 및 역량 유형의 전이 양상과 예측변인 분석: XGBoost와 잠재전이분석의 적용 - 저자 : 장은아(충남대학교) - 발행년 : 2025.02		
3	1%	[KCI 논문] KCI 논문 - 제목 : 법령 정보 확인 및 규제 판별에서의 생성형 AI 활용 가능성: ChatGPT 기반 탐색적 분석 - 저자 : 박정원(국립경국대학교) - 발행년 : 2025.06		
4	1%	[KCI 논문] KCI 논문 - 제목 : 딥러닝 알고리즘을 활용한 디스플레이 정전적 FAB 공정의 불량 탐지 및 원인 분석 - 저자 : 김서연(고려대학교) - 발행년 : 2022.04		
5	1%	[KCI 논문] KCI 논문 - 제목 : 해석가능 인공지능을 활용한 바이오화학 기술의 비즈니스 잠재성 평가 - 저자 : 이지호(건국대학교 산업공학과) - 발행년 : 2023.06		
6	1%	[KCI 논문] KCI 논문 - 제목 : 2차원 신체 관절 좌표를 활용한 피트니스 동작 평가 및 세분화 - 저자 : 박재욱(국민대학교 일반대학원 데이터사이언스학과) - 발행년 : 2024.12		
7	1%	[KCI 논문] www.kmis.or.kr - 제목 : STAGE-AI Metrics: 기업을 위한 생성형 AI 평가지표 설계 - 저자 : 한수민(성균관대학교) - 발행년 : 2025.02		
8	1%	[KCI 논문] KCI 논문 - 제목 : 고객의 속성별 중요도를 반영한 속성 기반 감성 분석(ABSA)을 이용한 레스토랑 추천 시스템 - 저자 : 우수현(경희대학교) - 발행년 : 2025.03		
9	1%	[KCI 논문] KCI 논문 - 제목 : Transformer-based Autoencoder와 FDD 손실 함수를 활용한 전류 센서의 비지도 학습 기반 이상 탐지 - 저자 : 박종호(국민대학교 데이터사이언스학과) - 발행년 : 2025.03		
10	1%	[KCI 논문] KCI 논문 - 제목 : 온라인 학습 데이터를 활용한 지식 추적 연구: 학생 능력 분포 차이가 모델의 성능에 미치는 영향을 중심으로 - 저자 : 권연하(건국대학교) - 발행년 : 2025.03		
11	1%	[KCI 논문] KCI 논문 - 제목 : PGA 투어 데이터를 활용한 골프 선수 경기력 예측 모델 개발 및 성능 분석 연구 - 저자 : 염두승(강원도립대학교) - 발행년 : 2025.09		
12	1%	[KCI 논문] KCI 논문 - 제목 : PageRank 특징을 활용한 RDP기반 내부전파경로 탐지 및 SHAP를 이용한 설명가능한 시스템 - 저자 : 윤지영(가천대학교) - 발행년 : 2021.08		

13	1%	[KCI 논문] KCI 논문 - 제목 : 설명 가능한 인공지능을 사용한 학습 데이터 전처리 방법 - 저자 : 이창홍(부산대학교) - 발행년 : 2023.02
14	1%	[KCI 논문] KCI 논문 - 제목 : 혼합물 실험계획법을 활용한 양상불 예측 모델 - 저자 : 권영식(공주대학교 산업공학과) - 발행년 : 2024.12
15	1%	[KCI 논문] KCI 논문 - 제목 : 캠핑 예약 플랫폼의 고객 행동 데이터 기반 고객 세그멘테이션 전략: RFM 분석 모델과 K-Means 클러스터링 기법 비교 를 기반으로 - 저자 : 문재형(한양대학교 기술경영전문대학원) - 발행년 : 2025.06

검사대상문서	비교대상문서
<div>문장유사율: 0%</div> <div><경영학연구 편집위원회 귀중> 1. 논문제목 (영문제목 포함) (국문) 데이터 중심 AI를 통한 구매 역동성과 설명력 규명: 마케팅 지식기반 특성공학으로 희소성 극복 (영문) Unveiling PurchaseDynamics and Explain ability via Data-Centric AI: Overcomin g Sparsity with Marketing Knowledge-Driven Feature Engineering 2.</div>	
<div>문장유사율: 0%</div> <div>투고분야 : Special Issue 1(Management of AI), 경영정보, 마케팅, 융합 3. 원고매수 : 총 53매 (참고문헌 포함) 4.</div>	
<div>문장유사율: 0%</div> <div>투고일자 : 2025.01.26 데이터 중심 AI를 통한 구매 역동성과 설명력 규명: 마케팅 지식기반 특성공학으로 희소성 극복 딥러닝이 갖는 지배적인 위치에도 불구하고, 데이터 가 희소하고정형 적이며 불균형이 심한 고객관계관리 분야에서의 효용성은 여전히 의문 의 여지가 있다.</div>	
<div>문장유사율: 0%</div> <div>본 연구는 무조건적인 딥러닝 도입을 지양하고, 이론 주도적 특성공학(Theory-Guided Feature Engineering) 프레임워크를 제안함으로써 기존 패러다임에 도전한다.</div>	
<div>문장유사율: 0%</div> <div>우리는 구매 주기, 시점 역동성, 마케팅 피로도, 경로 효율성 등 4가지 핵심 마케팅이론 을 모델이 학습 가능한 형태로 변환하여 귀납적 편향(Inductive Bias)으로 주입하였다.</div>	
<div>문장유사율: 0%</div> <div>극심한 희소성(전환율 0.12%)을 보이는 대규모 이커머스 데이터셋(REES46)을 통한 실증 분석 결과, 도메인 지식으로 강화된 트리 기반 모델(XGBoost)이 복잡한 딥러닝 모델을 성능과 효율성 모든 면에서 압도함을 확인하였다.</div>	
<div>문장유사율: 0%</div> <div>제안된 프레임워크는 베이스라인 대비 F1-score를 23.14% 향상되었으며, 동일 마케팅 예산하에서 전환 비용을 13.1% 절감할수 있음을 입증하였다.</div>	
<div>문장유사율: 0%</div> <div>또한, SHAP 분석을 통해 마케팅 피로도와 재구매 준비도가 구매 의사결정의 핵심 임계 점임을 규명함으로써 블랙박스 모델 대비 우수한 설명력을 제공하였다.</div>	
<div>문장유사율: 0%</div> <div>본 연구는 모델 중심에서 데이터 중심 AI로의 전환을 제안하며, 도메인 이론과 경량 알고리즘의 결합이 비즈니스의사결정을 위한 지속가능하고(Green AI) 설명 가능한 최적 해임을 시사한다.</div>	
<div>문장유사율: 0%</div> <div>주제어: 희소 데이터, 구매 예측, 도메인 지식 모델링, 머신러닝 vs 딥러닝, 설명가능한 AI Despite the ubiquity of deep learning, its efficacy in Customer Relationship M anagement, characterized by sparse, tabular, and highly imbalanced data, remai ns debatable.</div>	
<div>문장유사율: 0%</div> <div>This study challenges the indiscriminate application of deep learning by proposin g a Theory-Guided Feature Engineering framework.</div>	
<div>문장유사율: 0%</div> <div>We incorporate four key marketing theories: purchase cycle, temporal dynamics , marketing fatigue, and path efficiency, as inductive biases into the modeling pr ocess.</div>	
<div>문장유사율: 0%</div> <div>Empirical analysis using the extremely sparse REES46 dataset (0.12% conversio n rate) demonstrates that a tree-based model (XGBoost) enhanced with domain knowledge significantly outperforms complex deep learning architectures in both performance and computational efficiency.</div>	
<div>문장유사율: 0%</div> <div>The proposed framework achieved a 23.14% improvement in F1-score over bas elines and demonstrated a 13.1% reduction in conversion costs.</div>	

문장유사율: 0%

Furthermore, SHAP analysis identified marketing fatigue and repurchase readiness as critical decision-making thresholds, offering superior interpretability compared to black-box models.

문장유사율: 0%

Advocating for a shift from model-centric to Data-Centric AI, this research suggests that integrating domain theory with lightweight algorithms offers a sustainable (Green AI) and explainable optimal solution for business decision-making.

문장유사율: 0%

Keywords: Sparse Data, Purchase Prediction, Domain Knowledge Modeling, ML vs. DL, Explainable AI (XAI) I. 서론 오늘날의 고객은 즉각적으로 구매를 결정하지 않는다.

문장유사율: 0%

이메일을 열어보고, 며칠 뒤 모바일 푸시 알림을 확인하며, SMS를 수신한 뒤에 비로소 결제 버튼을 누른다.

출처 표시 문장

문장유사율: 0%

이처럼 산발적으로 흩어진 접점이 누적되어 하나의 복잡한 고객 여정(Customer Journey)이 완성된다(Harris et al., 2020).

문장유사율: 0%

기업 입장에서 이 파편화된 기록을 연결하여 누가 구매할지 예측하고, 한정된 마케팅 예산을 어디에 투입할지 결정하는 것은 수익성과 직결되는 지속 가능성을 위한 핵심 과제다.

문장유사율: 0%

그럼에도 불구하고 많은 기업 실무에서는 여전히 마지막 접점에만 기여도를 부여하거나, 임의의 가중치를 부여하는 단순 휴리스틱 모델에 의존하고 있어, 최종 구매에 이르기까지 고객이 거치는 선행 접점들의 기여도를 과소평가하고, 구매 의사결정의 인과관계를 지나치게 단순화하는 결과를 초래한다.

문장유사율: 0%

디지털 전환이 가속화됨에 따라 현대 기업은 전례 없는 양의 고객 상호작용 데이터를 축적하고 있다.

출처 표시 문장

문장유사율: 0%

고객 여정은 더 이상 선형적인 깔때기(Funnel) 구조가 아니라, 이메일 개봉, 앱 푸시 확인, 소셜미디어 클릭 등 다채널에서 발생하는 파편화된 접점들의 비선형적 집합체로 진화하였다(Vasileva, 2024).

출처 표시 문장

문장유사율: 0%

기업들은 이러한 페타바이트 규모의 로그 데이터가 고객 행동을 완벽하게 설명해줄 것이라는 데이터의 풍요를 기대하지만, 실제 분석 환경에서는 심각한 정보의 희소성(Sparsity) 문제에 직면한다(Jia et al., 2017).

문장유사율: 0%

일반적인 이커머스 환경에서 구매 전환율은 2~3% 수준에 불과하며, 이는 전체 데이터의 97% 이상이 구매와 직접적 관련이 없는 노이즈(Noise)이거나 단순 탐색 행동임을 의미한다.

출처 표시 문장

문장유사율: 0%

이러한 데이터의 희소성은 기계 학습 모델(Shin et al., 2022), 특히 최근 각광받는 딥러닝(Deep Learning) 모델의 학습을 저해하는 결정적 요인으로 작용한다.

출처 표시 문장

문장유사율: 0%

딥러닝 모델은 데이터 내의 복잡한 잠재 표현(Latent Representation)을 스스로 학습하기 위해 방대한 양의 밀도 높은 데이터를 필요로 하는데(Salakhutdinov, 2015), 고객 로그와 같은 스위스 치즈(Swiss Cheese) 형태의 구멍난 데이터 구조에서는 과적합(Overfitting)되거나 수렴에 실패할 가능성이 높기 때문이다(He et al., 2022).

출처 표시 문장

문장유사율: 0%

최근 학계와 산업계에서는 이미지 인식이나 자연어 처리(NLP) 분야에서 트랜스포머(Transformer)와 같은 딥러닝 아키텍처가 거둔 압도적 성과에 힘입어(Vaswani et al., 2017), 시계열 마케팅 데이터에도 LSTM(Long Short-Term Memory)이나 RNN(Recurrent Neural Network)을 적용하려는 시도가 주류를 이루고 있다(Du et al., 2019).

출처 표시 문장

문장유사율: 0%

이러한 접근은 원시 로그를 그대로 모델에 주입하여 인간의 개입 없이 특징을 추출하려는 표현 학습의 철학을 따른다(Bengio et al., 2012).

문장유사율: 0%

일부선행연구는 딥러닝이 기존 머신러닝을 더 좋다고 보고하며, "주요 해결책은 딥러닝"이라는 암묵적인 전제를 강화해왔다.

출처 표시 문장

문장유사율: 0%

하지만 여기에는 중요한 맹점이 존재한다. 딥러닝의 압도적 성과는 대개 데이터가 풍부하고 품질이 정제된 실험적 환경에서도출된 결과다(Sun et al., 2017). 반면, 실제 기업의 CRM 데이터는 고객의 구매 행동이 드문드문 발생하는 희소성이 높고, 이벤트 간의 시간 간격이 불규칙하며, 온라인과 오프라인 기록이 혼재되어 있다.

출처 표시 문장

문장유사율: 0%

이러한 희소한 데이터 환경에서는 수많은 파라미터를 가진 복잡한 딥러닝 구조가이론만큼의 성능 우위를 발휘하지 못할 가능성이 높다(Grinsztajn et al., 2022; Lindskog and Prehofer, 2023).

문장유사율: 0%

데이터의 밀도가 낮은 상황에서 무조건적인 고비용 모델 도입은 연산 비용과 마케팅 운영비용 측면에서 비효율적일 수 있다는 것이다(Farjon and Bar-Hillel, 2022).

문장유사율: 0%

2024년과 2025년에 발표된 다수의 대규모 벤치마크 연구들은 정형 데이터(Tabular Data) 영역에서 딥러닝의 우월성에 대해 심각한 의문을 제기하고 있다.

출처 표시 문장

문장유사율: 0%

이들 연구는 이질적인 변수들이 혼재된 정형 데이터 환경에서는 XGBoost나 CatBoost와 같은 트리 기반 앙상블 모델이 최신 딥러닝 모델보다 성능이 우수하거나 대등하면서도, 학습 비용은 현저히 낮음을 보고하고 있다(Hwang and Song, 2023; Wydmarski et al., 2023).

문장유사율: 0%

이는 마케팅 예측의 문제가 단순히 더 깊은 모델을 사용하는 것이 아니라, 데이터의 특성에 맞는 적정기술을 선택하는 문제임을 시사한다.

출처 표시 문장

문장유사율: 0%

특히 정형 데이터의 불연속적 속성 때문에의 딥러닝의 귀납적 편향이 맞지 않을 수 있다(Beyazit et al., 2023; Grinsztajn et al., 2022).

문장유사율: 0%

본 연구는 딥러닝 모델이 겪는 데이터 희소성 문제를 해결하기 위한 대안으로, 도메인 지식기반 특성공학(Feature Engineering)을 제안한다.

출처 표시 문장

문장유사율: 0%

마케팅 분야에는 Recency, Frequency, Monetary(RFM) 모형, 고객 생애가치(Customer Lifetime Value, CLV) 이론, 습관화 및 피로도(Fatigue) 이론 등 소비자 행동을 설명하는 검증된 이론들이 존재한다(Akter et al., 2025; Wathieu, 2004).

문장유사율: 0%

우리는 이러한 추상적 이론들을 모델이 학습 가능한 구체적인 수치형 변수로 변환함으로써, 모델에 강력한 귀납적 편향을 주입할 수 있다고 가정한다.

문장유사율: 0%

즉, 모델이 바닥부터 데이터의 패턴을 찾게 하는 대신, 시간이 지나면 반응이 감소한다(피로도)거나 월급날 근처에 구매 확률이 높다(시점 역동성)는 마케팅적 직관을 미리 계산된 형태로 제공하는 것이다.

문장유사율: 0%

이미 검증된 고객 행동이론이 존재하고 있고 마케터들에게는 익숙한 직관이자 지식일 수 있다.

문장유사율: 0%

그러나 기존 연구들은 이러한 도메인 지식을 단순히 현상을 요약하는 지표로만 썼을 뿐, 머신러닝 모델이 직접 학습할 수 있는 입력 피쳐로 체계적으로 재구성한 시도는 상대적으로 제한적이었다.

문장유사율: 0%

인간의 이론을 기계의 언어로 번역하여 주입한다면, 굳이 연산량이 많은 딥러닝 없이도 데이터가 희소한 상황에서도 높은 예측성능을 달성하고 모델이 신호와 소음을 구별하는데 결정적인 도움을 줄 수 있다.

인용 포함 문장

문장유사율: 0%

따라서 본 연구는 어떤 모델이 가장 고도화되고 복잡한지를 묻지 않는다. 대신 라는 실 "예측 정확도와 비용을 함께 고려할때, 어떤 조합이 가장 효율적인가?"용적 질문에 집중한다.

문장유사율: 0%

실제 대규모 이커머스 데이터셋(REES46)을 활용하여, 도메인 지식이 주입된 경량화된 모델이 원시 데이터 기반의 고비용 딥러닝 모델을 대체하거나 능가할 수 있는지를 실증적으로 규명하고자 한다.

문장유사율: 0%

구체적인 연구 질문은 다음과 같다. RQ1. [지식의 유용성] 마케팅도메인 지식(피로도, 구매 주기, 시점 역동성, 콘텐츠 효율성)을 반영한 파생변수는 단순 원시 로그만 사용했을 때보다 예측 성능을 유의미하게 향상시키는가? RQ2.

문장유사율: 0%

[해석 가능성] 4가지도메인 지식 축 중 예측력 개선에 가장 결정적인 역할을 하는 핵심 요인은 무엇이며, 이를 통해 실무자는 어떤 설명 가능한 마케팅 전략을 수립할 수 있는가? RQ3.

문장유사율: 0%

[알고리즘 비교] 희소성이 높은 CRM 데이터 환경에서, 도메인 지식 변수를 장착한 트리 기반 모델은 시계열 딥러닝 모델 대비 우월한 성능 및 효율성을 보이는가? 결론적으로 본 연구의 차별적 기여도는 다음과 같다.

문장유사율: 0%

첫째, 기존 연구들이고객 행동이론을 단순한 서술적 배경이나 사후 해석의 도구로만 활용한 것과 달리, 본 연구는 추상적인 마케팅이론을 머신러닝 모델이 학습 가능한 구체적인 수치형 벡터로 변환하는 체계적인 프레임워크를 제안한다.

문장유사율: 0%

이는 도메인 지식을 알고리즘의 귀납적 편향으로 주입하여, 데이터가 희소한 환경에서도 모델이 신호와 소음을 효과적으로 구분할 수 있게 한다.

문장유사율: 25%

둘째, 단순히 누가 구매할 것인지를 맞추는 블랙박스 예측을 넘어, **SHAP(SHapley Additive exPlanations) 분석을 통해** "왜 지금이 마케팅적기인가?"를 설명할 수 있는 모델을 구현하였다.

문장유사율: 0%

특히 마케팅 피로도와 재구매 준비도가 구매 확률에 미치는 비선형적 임계값을 규명함으로써, 실무자가 직관적으로 이해하고 즉각적인 마케팅 액션을 취할 수 있는 구체적인 근거를 제공한다.

문장유사율: 0%

셋째, 최근 학계의 Tabular Data에서의 딥러닝과 머신러닝 경쟁 논쟁에 기여하여, 구매 전환율이 극도로 낮은(0.12%) 실제 CRM 데이터 환경에서는 복잡한 딥러닝의 표현 학습보다 잘 설계된 도메인 특성을 장착한 경량 모델이 예측 성능과 일반화 능력 면에서 더 우월함을 실증적으로 규명한다.

문장유사율: 0%

이는 무조건적인 고비용 딥러닝 도입을 경계하고, 데이터 특성에 맞는 적정 기술의 중요성을 입증한다.

문장유사율: 44%

본 논문의 구성은 다음과 같다. 2장에서는 관련 선행연구를 고찰하고 본 연구의 지식 기반 변수 설계 근거를 제시한다.

문장유사율: 0%

3장에서는 연구실험 과정에서 사용된 데이터와 분석 알고리즘과 검증 방향을 상세하게 소개하고, 4장에서는 각 실험 결과와 해석을 소개한다.

KCI 논문 | 제목 : 설명가능한 AI ... | 저자 : 윤지운(한... | 발행년 : 2025.10

설명가능한 AI 기반 여자 기계체조 마루운동 심판 점수의 패턴 분석 및 편향 시각화 52
7 2. SHAP 기반 심판 점수 영향 요인 분석 **SHAP(SHapley Additive exPlanations)**
분석을 통해 실시점수(E)와 예술성점수(A)에 대한 변수 중요도를 평가한 결과, 심판ID
_학교급 변수가 두 점수 모두에서 모델 예측에 가장 큰 영향을 미치는 변수로 나타났다.

KCI 논문 | 제목 : Transformer-base... | 저자 : 박종호(국... | 발행년 : 2025.03

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 기존 연구에 대해서 서술한다.

KCI 논문 | 제목 : 설명 가능한 인공... | 저자 : 이창홍(부... | 발행년 : 2023.02

본 논문의 이후 **구성은 다음과 같다. 2장에서는 관련** 연구로 이미지 데이터를 사용한 인공지능 모델의 데이터 전처리의 중요성 및 기법 그리고 일반적인 설명 가능한 인공지능 알고리즘과 제안하는 시스템에서 사용된 알고리즘인 계층별 기여도 전파를 설명한다.

문장유사율: 0%

마지막으로 5장 토의에서 연구의 이론적 및 실무적 시사점과 한계를 논의한 후, 6장에서 연구결과를 요약하며 마무리한다.

문장유사율: 0%

II. 문헌연구 본 연구는 마케팅도메인 지식과 최신 기계학습 방법론의 융합을 통해 희소한 고객 데이터 환경에서의 예측 성능을 극대화하고자 한다.

문장유사율: 0%

이를 위해 (1) 고객 생애가치 및 행동이론의 진화, (2) 정형 데이터 환경에서의 딥러닝과 트리 기반 모델 간의 성능 논쟁, (3) 도메인 지식을 활용한 특성 공학의 이론적 타당성, (4) 설명 가능한 AI의 필요성을 중심으로 선행연구를 고찰한다.

출처 표시 문장

문장유사율: 0%

1. 고객 행동이론의 정량화와 진화 1) 고객 생애 가치와 RFM 프레임워크 고객 생애가치는 고객이기업과의 관계를 유지하는 전체 기간동안 창출할 것으로 기대되는 순현재 가치로 정의된다(Pfeifer et al., 2005).

문장유사율: 0%

이 개념은 단기적 거래 수익을 넘어 장기적 관계가치를 정량화 함으로써, 고객 획득 비용과 유지 투자의 효율성을 판단하는 핵심지표로 자리매김하였다.

출처 표시 문장

문장유사율: 0%

CLV는 1980년대 직접 마케팅에서 출발하여, 1990년대 데이터베이스 마케팅, 2000년대 e-commerce 및 구독 경제, 그리고 2020년대 옴니채널과 실시간 개인화 환경으로 그 적용범위가 확장되어 왔으며(Berger and Nasr, 1998), 방법론 또한 집계형 RFM에서 확률 모형으로, 그리고 머신러닝 융합이라는 진화 궤적을 따르고 있다.

문장유사율: 0%

고객가치를 평가하는 가장 고전적이면서도 강력한 프레임워크는 RFM 모형이다. 초기 연구들은 RFM 지표가 미래 수익성과 강한 상관관계를 가짐을 실증하였다.

문장유사율: 0%

그러나 단순한 집계 방식의 RFM은 고객의 이탈(Churn) 시점을 명시적으로 알 수 없는 비계약적 환경에서의 행동을 설명하는데 한계를 보였다.

문장유사율: 0%

이를 극복하기 위해 Schmittlein et al. (1987)은 Pareto/NBD 모형을, Fader (2005)은 이를 개선한 BG/NBD 모형을 제안하며 Buy Til You Die (BTYD) 프레임워크를 확립하였다.

문장유사율: 0%

이들은 고객의 구매 이력을 확률과정(Stochastic Process)으로 모델링하여 잠재적인 이탈 확률을 추정하였다.

문장유사율: 0%

특히 Fader et al. (2005)이 제기한 빈도의 역설(Paradox of Frequency)은 구매 빈도가 높더라도 최근성(Recency)이 개인의 평균 주기보다 길어지면이탈 확률이 급격히 증가한다는 현상을 언급하였다.

문장유사율: 0%

본 연구에서 제안하는 재구매 준비도 변수의 이론적 토대가 된다. 하지만 이러한 확률 모형들은 데이터의 이질성(Heterogeneity)을 설명하기 위해 엄격한 분포 가정을 필요로 하며, 마케팅 믹스 변수나 시변 공변량(Time-varying Covariates)을 유연하게 통합하기 어렵다는 구조적 한계를 갖는다. 최근 연구들은 CLV 추정의 정확화를 위해 기계학습기법과의 융합을 시도하고 있다.

출처 표시 문장

문장유사율: 0%

예를 들어, 생존 분석(Survival Analysis)과 순환 신경망을 결합하여 고객의 이탈 위험과 미래 거래가치를 동시에 예측하는 방법론이 제안되었으며 (Chamberlain et al., 2017) 이는 전통적인 확률 모형의 한계를 보완하는 것으로 평가된다.

문장유사율: 0%

그러나 이러한 접근법은 대부분 모델의 복잡성 증대에 초점을 맞추고 있으며, 도메인 지식을 명시적으로 통합하는 시도는 여전히 제한적이다.

문장유사율: 0%

본 연구는 이러한 간극을 메우고자, Fader et al. (2005)의 이론적 프레임워크를 기반으로 구매 주기의 개인별 이질성을 정규분포로 모델링하고, 현재 시점이 고객의 평균 재구매 주기에 얼마나 근접했는지를 확률 밀도 함수로 정량화하는 재구매 준비도 지표를 제안한다.

문장유사율: 0%

2) 시계열적 역동성과 피로도 이론 소비자의 반응은 물리적 시간의 흐름과 마케팅 자극의 강도에 따라 비선형적으로 변화한다.

문장유사율: 0%

Koren (2009)의 연구는 사용자 선호가 고정된 것이 아니라 시간에 따라 표류한다는 사실을 규명하며 시점 역동성(Temporal Dynamics) 모델링의 중요성을 강조하였다.

문장유사율: 0%

그는 추천시스템 분야에서 협업 필터링(Collaborative Filtering) 모델에 시간의존적 편향(Time-varying Bias)을 도입함으로써, 예측 정확도를 유의미하게 개선하였다.

문장유사율: 0%

이러한 연구는 시간적 맥락이 예측 모델에 필수적으로 통합되어야 함을 보여주는 이론적 토대를 제공한다.

출처 표시 문장

문장유사율: 0%

최근 이커머스 환경에서는 행동경제학 및 마케팅 분야의 연구를 기반으로 특정 요일이나 급여일 전후로 구매 확률이 변동하는 캘린더 효과(Calendar Effect)가 주요한 예측 변수로 다루며 특정 시점에 소비자의 구매력과 지출 성향이 일시적으로 변동함을 실증하였다(Olafsson and Pagel, 2018).

문장유사율: 0%

이러한 발견은 본 연구에서 제안하는 급여일 근접도 및 분기말 근접도 변수의 이론적 정당성을 뒷받침한다.

문장유사율: 0%

특히 2020년 이후 팬데믹, 인플레이션 등 거시 충격이 소비패턴을 급변시키면서, 시간 의존적 모델링의 실무적 중요성은 더욱 부각되고 있다.

출처 표시 문장

문장유사율: 0%

또한, 소비자의 활동 리듬과 미디어 소비패턴의 관계를 규명한 연구들은, 메시지 전달 시점의 최적화가 단순히 메시지 도달 확률을 높이는 것을 넘어 인지적 수용성을 극대화한다는 점을 강조한다 (Althoff et al., 2017).

문장유사율: 0%

본 연구는 이러한 통찰을 반영하여, 원형 통계(Circular Statistics) 기법을 활용해 개별 고객의 과거 반응 시점을 단위 원 상의 벡터로 매핑하고, 현재 발송 시점과의 코사인 거리를 산출함으로써 선호 시간대와 얼마나 일치하는지 정도를 변수를 구성하였다.

문장유사율: 0%

한편, 정보처리 이론(Information Processing Theory)과 습관화 이론은 반복적인 마케팅 자극이 소비자의 주의를 감소시키고 반응 역치를 높인다고 설명한다.

문장유사율: 0%

이는 마케팅 피로도(Marketing Fatigue)로 개념화되며, 과도한 메시지 노출이 오히려 클릭률(CTR)과 전환율을 저하시키는 역U자형 관계를 형성함을 시사한다.

출처 표시 문장

문장유사율: 0%

심리학 분야에서는 이를 Weber-Fechner 법칙의 맥락에서 설명하는데, 동일한 자극의 반복은 감각 역치를 상승시켜 추가적인 자극에 대한 변별력을 저하시킨다(Ranganathan and Ganapathy, 2002).

문장유사율: 0%

마케팅 맥락에서 이는 광고 회피나 이메일 무시와 같은 형태로 발현된다. 최근 추천시스템 분야의 연구는 이러한 피로도 개념을 정보 필터링의 맥락으로 확장하였다.

문장유사율: 0%

Ziegler et al. (2005)은 추천 목록 내 항목 간 유사도가 높을수록 사용자의 만족도가 저하됨을 실증하였으며, 이를 한계 효용 체감의 법칙(Law of Diminishing Marginal Returns)으로 해석하였다.

문장유사율: 0%

이들의 연구는 공간적 차원에서의 유사성이 효용 감소를 유발한다는 점을 보였으나, 본 연구는 이를 시간적 차원으로 확장하여, 동일 채널을 통한 반복적 메시지 노출이 시간의 흐름에 따라 피로도를 누적시킨다는 가설을 제안한다.

문장유사율: 0%

최근 연구들은 이러한 피로도를 모델링 하기 위해 지수 감쇠(Exponential Decay) 함수 등을 도입 하고 있으나, 이를 머신러닝의 입력 변수로 동적으로 구현하여 최적화한 시도는 여전히 부족하다.

문장유사율: 0%

본 연구는 이러한 이론적 배경을 바탕으로, 채널별 반감기를 고려한 누적 피로도 함수를 제안하여 모델에 내재된 과적합 위험을 제어하고자 한다.

출처 표시 문장

문장유사율: 0%

실제 각 채널의 피로도 반감기는 매체의 침해성과 고객 반응 임계점에 대한 선행 연구를 기반으로 차등 설정되었다(Attentive, 2025; Pilot and Rello, 2017).

문장유사율: 0%

특히, 최근 7일간 노출 빈도와 마지막 노출 후 경과 시간을 결합한 복합 감쇠 함수를 적용하였으며 여기서 핵심 하이퍼 파라미터인 (피로 강도)와 (반감기)는 채널별 하이퍼 파라미터는 다음과 같은 실증적 근거에 따라 정의된다: SMS() < 푸시() < 이메일().

출처 표시 문장

문장유사율: 0%

첫째, SMS는 개인 공간 침해에 따른 높은 저항성(Attentive, 2025; Pilot and Rello, 2017)과 구매의도 소멸의 임계점(GrowthSuite, 2026; Rejoinder, 2025)을 반영하여 가장 낮은 허용 빈도()와 가장 긴 반감기()를 설정하였다.

문장유사율: 0%

이는 장바구니 포기 후 구매 의도가 급격히 소멸되는 시간을 역으로 적용하여, 고객의 부정적 피로감 역시 해당 시간 동안 잔존한다고 보수적으로 가정한 결과이다.

문장유사율: 0%

둘째, 모바일 푸시는 화면 점유형 알림이나, 스와이프로 즉시 제거되는 매체의 높은 휘발성과 빠른 심리적 잔상 소멸을 고려하여 중간 수준의 빈도()와 가장 짧은 반감기()를 적용하였다.

문장유사율: 0%

이는 알림이 쌓이는 이메일이나 SMS와 달리 피로도가 빠르게 회복되는 매체 특성을 반영한 것이다.

출처 표시 문장

문장유사율: 0%

셋째, 이메일은 비동기적 채널로서 피로 민감도가 낮아 가장 높은 임계값()을 갖는다. 이는 주 2회 발송이 최적이라는 대규모 분석 결과(MailerLite, 2025)에 근거하며, (은 이러한 재발송 주기를 수학적으로 지지한다.

출처 표시 문장

문장유사율: 0%

이러한 채널별 차별화는 단순히 경험적 관찰에 그치지 않고, 매체 풍부성 이론의 연장선상에서 이론적으로 정당화될 수 있다(Daft and Lengel, 1986).

문장유사율: 0%

3) 마케팅 콘텐츠 신규성 및 경로 효율성이론 마케팅 메시지의 효과는 단순히 노출 빈도나 타이밍에만 의존하지 않으며, 메시지 자체의 내용적 특성과 고객의 과거 경험 간의 상호 작용에 의해 조절된다.

문장유사율: 0%

이러한 관점은 크게 두가지 이론적 흐름으로 구분된다: (1) 콘텐츠 신규성이주의 환기와 정보가치에 미치는 영향, (2) 설득 경로의 구조적 효율성이 전환 확률에 미치는 영향이다.

문장유사율: 0%

콘텐츠 신규성의 중요성은 정보 이론과 심리학의 새로운 선행 연구로부터 도출된다.

문장유사율: 0%

Shannon (1948)의 정보 엔트로피 개념에 따르면, 예측 가능한 메시지는 정보량이 낮으며 수신자에게 전달하는가치가 제한적이다.

문장유사율: 0%

반면, 새로운 정보는 높은 엔트로피를 가지며 주의를 환기시키는 효과가 크다. 마케팅 맥락에서 이는 동일한 주제나 제안의 반복적 노출이 정보의 한계 효용을 감소시킨다는 것을 의미한다.

문장유사율: 0%

Ziegler et al. (2005)는 주제다 양화를 통해 완화할 수 있음을 보였다. 본 연구는 이러한 공간적 유사성 개념을 시간적 차원으로 전환하여, 특정 주제에 대한 최근 노출 빈도와 마지막 노출 이후 경과 시간을 결합한 복합 감쇠 함수를 제안한다.

문장유사율: 0%

이 함수는 주제 과포화상태를 정량화 함으로써, 마케터가 메시지 다양성을 전략적으로 관리할 수 있는 지표를 제공한다.

출처 표시 문장

문장유사율: 0%

설득 경로 효율성은 고객 여정 연구의 핵심 주제이다. 전통적인 AIDA(Attention, Interest, Desire, Action) 모델이나 깔때기 프레임워크는 선형적 단계들이 정하지만, 현대의 옴니채널 환경에서 고객 여정은 고도로 비선형 적이며 개인별로이질적이다(Lemon and Verhoef, 2016).

문장유사율: 0%

디지털 전환 이후 전통적 선형 모델이 비선형 및 다중 접점 고객 여정 모델로 대체되는 패러다임 전환이 진행 중이며, 이러한 복잡성 속에서 특정 채널 조합이 전환으로 이어지는 성공 경로를 식별하는 것은 마케팅 효율성 극대화의 핵심과제로 부상하였다.

문장유사율: 0%

Anderl et al. (2016) 은 그래프 기반 마코프 모델(Markov Graph Model)을 활용하여 온라인 광고 채널간의 기여도를 정량화 하였다.

문장유사율: 0%

전통적인 Last-Click 귀속 모델이 초기 접점의 기여를 과소평가 한다는 점을 실증하였으며, 채널간 상호작용을 고려한 속성 분석의 중요성을 강조하였다.

문장유사율: 0%

특히 그들이 발견한 경로의존성 현상, 즉 동일한 채널이라도 선행 접점의 구성에 따라 효과가 달라진다는 결과는, 본 연구에서 제안하는 성공 경로 일치도 변수의 이론적 근거를 제공한다.

문장유사율: 0%

종합하면, 본 연구의 콘텐츠 및 경로 효율성 변수군은 고객의 과거 경험을 학습 데이터로 활용하여, 미래 메시지의 설득력을 구조적으로 예측하는 메커니즘을 구현한다.

문장유사율: 0%

이는 단순히 무엇을 보낼 것인가를 넘어 어떻게 보낼 것인가의 전략적 최적화를 가능케 한다.

문장유사율: 0%

2. 지식기반 특성 공학을 통한 이론의 귀납적 편향 주입 연구 특성공학은 단순히 데이터를 전처리하는 기술적 과정을 넘어, 문제해결을 위한 인간의 지식을 모델에 전달하는 지식 주입과 정으로 재정의 되어야 한다.

문장유사율: 0%

마케팅 분야에서도메인 지식을 데이터 분석에 통합하려는 시도는 오랜 역사를 가지고 있으나, 최근 머신러닝 및 인공지능 기술의 발전과 함께 그 양상이 질적으로 변화하고 있다.

문장유사율: 0%

전통적으로 도메인 지식은 모델링 이전 단계에서 변수 선택이나 데이터 전처리의 지점으로 활용되었으나, 최근 연구들은 이를 모델의 구조적 제약이나 학습 과정의 귀납적 편향으로 명시적으로 주입하려는 시도를 보이고 있다.

문장유사율: 0%

이러한 흐름은 크게 세 단계로 구분될 수 있다: (1) 초기 단계에서는 이론적 개념을 수작업으로 수식화 하여 변수로 변환하는 전통적 피쳐 엔지니어링이 주류를 이루었고, (2) 확장 단계에서는 AutoML과 같은 자동화 도구가 피쳐 생성 및 선택과정을 효율화 하였으며, (3) 최근에는 거대언어모델(LLM)을 활용하여 이러한 지식들을 반영하는 자동화 시도가 등장하고 있다.

문장유사율: 0%

이는 도메인 지식이 인간 전문가의 전유물이 아니라, 방대한 텍스트 데이터에서 추출 가능한 일반화된 지식임을 시사한다.

문장유사율: 0%

본 연구는 이러한 흐름의 연장선상에서, 도메인 지식을 체계적으로 변수화하는 프레임워크를 제시한다.

문장유사율: 0%

초기 단계는 모델에 도메인 이론을 명시적인 제약 조건이나 파생 변수의 형태로 주입하는 방식으로, 순수 데이터 기반 접근법이 데이터 부족이나 노이즈로 인해 실패할 때, 이론적 지식이 모델의 탐색 공간을 효율적으로 제한하는 역할을 할 수 있다.

문장유사율: 23%

Fader et al. (2005)은 RFM 이론을 Pareto/NBD 확률 모형과 연결함으로써, 고객의 구매 이력이라는 관찰 데이터를 잠재적 행동 성향의 충분통계량으로 변환하였다.

KCI 논문 | 제목 : 캠핑 예약 플랫폼... | 저자 : 문재형(한... | 발행년 : 2025.06

Fader et al. (2005)은 Frequency와 Monetary 간의 높은 상관관계가 발견되어 변수 간 중복성이 실증적으로 확인한 바에 있어 RFM 변수간 상호 독립성이 보장되지 않는다는 한계가 존재한다.

문장유사율: 0%

이는 단순히 과거 데이터를 집계하는 것을 넘어, 확률 모형의 파라미터 추정을 통해 미래 행동을 예측 가능한 형태로 정제한다는 점에서 의의가 있다.

문장유사율: 0%

그러나 이러한 접근법은 특정 확률 분포에 대한 강한 가정을 요구하며, 실제 데이터가 가정을 위배할 경우 추정치의 신뢰성이 저하된다는 한계를 갖는다. Kumar et al. (2019)은 고객의 소셜 네트워크 구조와 과거 구매 이력을 결합하여 영향력 점수라는 파생 변수를 생성하였으며, 이를 추천 시스템에 투입함으로써 예측 정확도를 개선하였다.

문장유사율: 0%

이들의 접근법은 사회적 영향력이론을 네트워크 중심성 지표로 조작화한 것으로 볼 수 있으며, 본 연구의 방법론적 지향과 유사한 맥락을 공유한다.

문장유사율: 0%

한편으로, 피로도 및 습관화 모델링은 디지털 마케팅의 맥락에서 특히 중요성이 부각되고 있다.

문장유사율: 0%

Lambrecht and Tucker (2013)는 리타겟팅 광고의 빈도가 과도할 경우 오히려 전환율이 감소하는 역U자형 관계를 실증하였으며, 이는 피로도 현상의 경험적 증거로 해석된다.

문장유사율: 0%

그러나 이들의 연구는 집계 수준의 분석에 한정되어 있어, 개별 고객의 피로도를 동적으로 추적하는 메커니즘을 제공하지 못한다.

문장유사율: 0%

본 연구는 지수 감쇠 모델을 활용하여 각 고객의 채널별 누적 피로도를 실시간으로 계산하며, 채널의 집해성에 따라 반감기를 차등화 함으로써 이론적 정교성을 높였다.

출처 표시 문장

문장유사율: 0%

최근에는 거대언어모델을 활용한 연구들이 위치한다. 전자상거래 분야에서는 LLM을 이용해 상품설명 텍스트로부터 자동으로 마케팅 관련 특성(예: 계절성, 가격 민감도)을 추출하고, 이를 수요예측 모델의 입력으로 활용하는 프레임워크를 제안하였다(Zhang et al., 2024).

문장유사율: 0%

이러한 접근법은 도메인 지식이 인간 전문가의 암묵지가 아니라, 방대한 텍스트 데이터로부터 추출 가능한 명시적 지식으로 전환될 수 있음을 시사한다.

문장유사율: 0%

그러나 LLM 기반 접근법은 생성된 특성의 해석 가능성과 인과적 타당성이 보장되지 않는다는 근본적 한계를 갖는다. Wang et al. (2019)은 상품 속성, 사용자 프로필, 맥락 정보를 다차원 그래프로 구조화하고, 그래프 신경망을 통해 임베딩을 학습함으로써 추천 성능을 개선하였다.

문장유사율: 0%

이들의 접근법은 도메인 지식을 관계형 구조로 명시화한다는 점에서 본 연구와 유사하나, 그래프 학습 과정에서 원래의 의미적 관계가 잠재 공간으로 압축되어 해석 가능성이 저하된다는 차이가 있다.

문장유사율: 0%

결국, 이러한 도메인 특성 공학이 LLM이나 복잡한 딥러닝 아키텍처보다 실무적 예측 성능 향상에 크게 기여함을 보고하고 있다.

문장유사율: 0%

3. 정형 데이터 분석에서의 알고리즘 논쟁 최근 학계, 특히 2024년과 2025년의 주요 화두 중 하나는 정형 데이터 영역에서 딥러닝이 과연 전통적인 트리 기반 모델을 능가하는가에 대한 논쟁이다

문장유사율: 0%

(Gorishniy et al., 2024; Grinsztajn et al., 2022; Somvanshi et al., 2024). 이미지(CNN)나 텍스트(Transformer) 분야에서 딥러닝이 거둔 압도적 성과에 힘입어, 마케팅 분야에서도 고객 행동 로그를 시계열 데이터로 간주하고 LSTM이나 TabNet, TabTransformer와 같은 딥러닝 아키텍처를 적용하려는 시도가 급증했다

문장유사율: 0%

((Arik and Pfister, 2021; Chen et al., 2019; Huang et al., 2020). 이들은 딥러닝이 수작업 특징 추출 없이도 원시 데이터로부터 고차원의 잠재 표현을 학습할 수 있다는 End-to-End 학습의 장점을 강조한다(Gorishniy et al., 2024; Shwartz-Ziv and Armon, 2022).

문장유사율: 0%

마케팅 분야에서 머신러닝과 딥러닝 기법의 활용은 지난 수년간 크게 증가하였으며, 그 적용 범위는 고객 세분화, 이탈 예측, 추천시스템, 가격 최적화 등 다방면에 걸쳐 있다.

문장유사율: 0%

그러나 모델의 예측 성능 향상에도 불구하고, 마케팅의사결정의 맥락에서 요구되는 해석 가능성과 인과적 추론 능력은 여전히 충분히 확보되지 않았다는 비판이 제기되고 있다 (Poursabzi-Sangdeh et al., 2021). 그 중 이탈 예측 및 고객 생애가치 추정에는 머신러닝 응용의 전통적 영역이다.

문장유사율: 0%

Verbeke et al. (2012)은 로지스틱 회귀, 의사결정나무, 서포트 벡터 머신 등 다양한 알고리즘을 비교하였으며, 앙상블 기법(Ensemble Methods)이 단일 모델 대비 우수한 성능을 보임을 실증하였다.

문장유사율: 0%

이후 XGBoost와 LightGBM과 같은 그라디언트 부스팅 계열 알고리즘이 Kaggle 경진대회 등에서 압도적 성능을 보이며 실무 표준으로 자리잡았다. 그러나 이들 모델은 주로 정형 데이터에 최적화 되어 있으며, 시계열적 의존성이나 순차적 패턴을 명시적으로 학습하는 데는 구조적 한계를 갖는다. 딥러닝의 도입은 이러한 한계를 극복하려는 시도로 이해될 수 있다.

문장유사율: 0%

순환 신경망과 그 변형인 LSTM은 고객 행동의 시계열적 패턴을 학습하는데 활용되어 왔다.

문장유사율: 0%

예를 들어, Jannach and Ludewig (2017)은 세션 기반 추천시스템에 RNN을 적용하여, 사용자의 단기 의도를 포착하는데 성공하였다.

출처 표시 문장

문장유사율: 0%

이후 Self-Attention 메커니즘을 활용한 Transformer 기반 모델들이 제안되었으며(Kang and McAuley, 2018), 이들은 장거리 의존성(Long-range Dependency)을 효과적으로 포착할 수 있다는 장점을 보였다.

문장유사율: 0%

He et al. (2020)은 상품 이미지와 사용자 리뷰 텍스트를 결합한 CNN-LSTM 하이브리드 모델을 제안하였으며, 이는 단일 모달리티 모델 대비 추천 정확도를 개선하였다.

문장유사율: 0%

그러나 이러한 접근법은 본 연구와 같이 구조화된 로그 데이터만 가용한 환경에는 적용이 제한적이다.

문장유사율: 0%

최근의 벤치마크 연구들은 정형 데이터 환경에서 딥러닝의 우월성에 대해 근본적인 의문을 제기하고 있다.

문장유사율: 0%

Grinsztajn et al. (2022)은 45개의 데이터셋을 대상으로한 대규모 실험을 통해, 트리 기반 모델이 ResNet, Transformer 등의 딥러닝 모델보다 정형 데이터에서 일관되게 우수한 성능을 보임을 실증하였다.

문장유사율: 0%

이들은 그 원인을 다음과 같이 분석한다: (1) 정형 데이터는 이미지나 텍스트와 달리 변수 간 평활성(Smoothness)이 보장되지 않으며, (2) 트리 모델의 축 정렬 분할(Axis-aligned Splits) 방식이 이산적 또는 국소적 패턴을 포착하는데 더 적합하다.

문장유사율: 0%

이러한 발견은 본 연구가 최소한 CRM 데이터 환경에서 트리 기반 모델을 기존 모델로 선택한 이론적 근거를 제공한다.

문장유사율: 0%

Shwartz-Ziv and Armon (2022)의 연구는 한걸음 더 나아가, 정형 데이터에서 딥러닝이 실패하는 구조적 원인을 규명하였다.

문장유사율: 0%

그들은 다수의 딥러닝 아키텍처와 정형 데이터 전용 모델을 비교한 결과, XGBoost가 가장 높은 성능을 보였으며, 딥러닝 모델 중에서는 ResNet과 FT-Transformer가 상대적으로 양호한 성능을 보였으나 여전히 XGBoost를 능가하지 못했다.

문장유사율: 0%

Gorishniy et al. (2021)은 Tabular 데이터 전용 딥러닝 아키텍처인 FT-Transformer를 제안하였으며, 이는 범주형 변수와 수치형 변수를 모두 임베딩(Embedding)하여 Transformer에 입력하는 방식이다.

문장유사율: 0%

이들은 다수의 데이터셋에서 XGBoost와 경쟁적인 성능을 달성하였으나, 학습시간과 하이퍼 파라미터 민감도 측면에서는 여전히 트리 모델이 더 효율적임을 인정하였다.

문장유사율: 0%

이러한 결과는 딥러닝이 정형 데이터에서 절대적 우위를 점하기보다는, 특정 조건에서만 그 효용이 발휘됨을 시사한다.

문장유사율: 0%

Chen et al. (2023)의 최근 연구는 이러한 논쟁에 균형 잡힌 시각을 제공한다. 그들은 176개의 데이터셋을 대상으로한 메타 분석을 통해, 데이터셋의 특성(샘플 수, 변수 수, 클래스 불균형 등)에 따라 최적 모델이 달라짐을 발견하였다.

문장유사율: 0%

특히 소규모 데이터셋에서는 트리 모델이 우위를 보이며, 대규모 데이터셋에서는 딥러닝 모델이 경쟁력을 갖는다는 결과를 제시하였다.

문장유사율: 0%

본 연구의 데이터셋은 이론적으로 딥러닝의 효용이 발현될 수 있는 규모이나, 극심한 클래스 불균형과 변수 간 이질성이 높다는 특성을 고려할 때, 트리 모델이 더 적합할 것으로 예상된다.

출처 표시 문장

문장유사율: 0%

최근 2025년에 발표된 연구들 또한 딥러닝 모델이 하이퍼 파라미터 튜닝에 극도로 민감하며, 학습 비용 대비 성능 효율성이 떨어진다는 점을 지적한다(Attari and Arroyave, 2025; Hollmann et al., 2025).

출처 표시 문장

문장유사율: 0%

이러한 현상이 발생하는 핵심 원인은 데이터의 희소성과 모델의 귀납적 편향의 불일치에 있다(Beyazit et al., 2023; Grinsztajn et al., 2022).

문장유사율: 0%

최근에는 Green AI 담론과 EU AI Act 등 규제 환경의 변화로 인해, 단순한 예측 성능을 넘어 모델의 복잡성 대비효율, 해석 가능성, 그리고 운영 지속 가능성이 재조명되고 있다.

문장유사율: 0%

이러한 흐름은 무조건적인 딥러닝 도입보다는 문제 특성에 맞는 적정기술 선택의 중요성을 강조하며, 본 연구가 트리 기반 모델과 명시적 피처 엔지니어링을 채택한 배경과 맥을 같이한다.

문장유사율: 0%

특히 정형 데이터 환경에서는 트리 기반 모델의 효율성과 강건성이 재평가되고 있으며, 딥러닝은 비정형 데이터나 대규모 및 고차원데이터에서 그 진가를 발휘한다는 인식이 점차 공유되고 있다.

문장유사율: 0%

따라서 본 연구는 무조건적인 딥러닝 도입을 지양하고, 데이터의 특성에 적합한 모델을 베이스 라인으로 선정하되, 딥러닝의 한계를 극복하기 위한 새로운 접근법을 모색한다.

문장유사율: 0%

4. 설명가능한 AI 예측 모델의 성능만큼이나 중요한 과제는 모델의 결과를 마케터가 신뢰하고 의사결정에 활용할수 있는가이다.

출처 표시 문장

문장유사율: 0%

2025년 현재, 단순히 학술적 관심사를 넘어, 유럽의 AI 법안(EU AI Act) 등 규제 환경의 변화로 인해, 예측 모델의 투명성은 선택이 아닌 필수가 되었다(EU, 2024; Hacker, 2024).

문장유사율: 0%

딥러닝 모델, 특히 순환 신경망은 내부 작동원리를 파악하기 어려운 블랙박스로 간주된다.

출처 표시 문장

문장유사율: 0%

반면, 트리 기반 모델은 SHAP 값과 결합될 때 각 변수가 예측에 기여한 정도를 게임 이론적으로 완벽하게 분해할 수 있다(Lundberg et al., 2020; Lundberg and Lee, 2017).

출처 표시 문장

문장유사율: 0%

이는 마케터에게 "왜 이 고객에게 추천해야 하는가?"에 대한 근거를 제공하여, AI 도입의 가장 큰 장벽인 신뢰 문제를 해결한다(Glikson and Woolley, 2020; McKinsey & Company, 2024).

문장유사율: 0%

트리 기반 모델의 내재적 해석 가능성은 오랫동안 그 장점으로 인정되어 왔다. 각 분기 노드에서의 규칙은 인간이 직관적으로 이해할 수 있는 IF-THEN 형식을 따르며, 변수 중요도(Feature Importance)는 각 변수가 전체 예측에 기여한 정도를 정량화한다.

문장유사율: 0%

Random Forest나 XGBoost와 같은 앙상블 모델의 경우, 개별 트리의 복잡성으로 인해 완전한 투명성은 보장되지 않으나, 변수 중요도의 집계를 통해 전역적 해석은 여전히 가능하다.

문장유사율: 27%

LIME(Local Interpretable Model-agnostic Explanations)은 Ribeiro et al. (2016)이 제안한 사후 해석 기법으로, 예측하려는 관측치 주변의 국소 영역에서 선택 모델을 근사함으로써 설명을 생성한다.

KCI 논문 | 제목: 설명가능한 AI ... | 저자: 윤지운(한... | 발행년: 2025.10

대표적인 기법인 SHAP(SHapley Additive exPlanations)과 LIME(Local Interpretable Model-agnostic Explanations)은 모델 예측 결과와 변수 간 상호작용을 설명하는데 효과적으로 활용된다(Lundberg & Lee, 2017; Ribeiro et al. 2016).

KCI 논문 | 제목: STAGE-AI Metr... | 저자: 한수민(성... | 발행년: 2025.02

LIME(Local Interpretable Model-Agnostic Explanations)(Ribeiro et al. 2016)은 AI 모델의 예측 결과를 설명하기 위한 해석 가능성 평가 벤치마크로, 복잡한 모델의 의사결정 과정을 단순화해 각 특성이 예측에 미치는 영향력을 정량적으로 평가한다.

KCI 논문 | 제목: 해석가능 인공지능... | 저자: 이지호(건... | 발행년: 2023.06

대표적으로, 모델의 국소적 예측 결과값에 대한 해석을 수행하는 LIME(Local Interpretable Model-agnostic Explanations) 알고리즘과, Anchor 알고리즘이 개발되었으며(Ribeiro et al. 2016; Ribeiro et al. 2018), 모델의 입력 값의 변화가 모델 예측치에 미치는 영향 정도인 Shapley value를 활용하는 SHAP (SHapley Additive exPlanation) 알고리즘이 개발되었다(Lundberg and Lee, 2017).

문장유사율: 0%

LIME은 모델에 구애받지 않는 특성 덕분에 딥러닝 모델에도 적용 가능하나, 국소 근사의 안정성이 보장되지 않으며 반복 실행시 설명이 변동할 수 있다는 한계가 있다(Alvarez-Melis and Jaakkola, 2018). 본 연구는 SHAP의 공리적 안정성을 우선시하여 LIME은 보조적으로만 활용한다.

출처 표시 문장

문장유사율: 0%

트리 모델과 SHAP의 결합은 특히 강력한 설명력을 제공한다. TreeSHAP 알고리즘(Lundberg et al., 2020)은 트리 앙상블 모델에 대해 정확한 Shapley Value를 다항 시간 안에 계산할 수 있으며, 이는 모델의 예측을 개별 관측치 수준에서 완전히 분해할 수 있음을 의미한다.

문장유사율: 0%

본 연구는 이러한 기법을 활용하여, 최종 모델이 어떤 변수들의 조합을 통해 고성능을 달성했는지를 규명하며, 특히 제한된 도메인 지식 변수들이 기존 베이스라인 변수 대비 어떤 차별적 설명력을 갖는지를 비교 분석한다.

인용 포함 문장

문장유사율: 0%

III. 데이터 및 방법론 1. 데이터 수집 및 특성 본 연구는 제안하는 도메인 지식기반 예측 방법론의 실증적 유효성과 실무적 재현성을 확보하기 위해, 대규모 이커머스 환경의 실제 로그를 담고 있는 Kaggle의 데이터셋(REES46)을 분석 대상으로 선정하였다.
"E-commerce multichannel direct messaging 2021-2023"

문장유사율: 0%

이 데이터는 2021년부터 2023년까지 2년간 중형이 커머스 플랫폼에서 수집된 고객관계관리(CRM) 기반의 마케팅 메시징 로그로, 개인화된 이메일, 웹이나 모바일 푸시 등의 프로모션과 리텐션(Retention) 활동을 포함한 실제 비즈니스 상 전자제품, 화장품, 다품목 스토어 등 다양한 카테고리의 운영 기록을 익명화하여 구성한 것이다.

문장유사율: 0%

구체적으로 해당 데이터셋은 총 1억 7,200만 건의 방대한 메시지 발송 로그를 포함하고 있으며, 약 185만명의 사용자를 대상으로 총 1,907개의 마케팅 캠페인의 실행 이력이 포함되어 있다.

문장유사율: 0%

분석에 활용된 원시 데이터(Raw Data)는 고객의 메시지 수신 및 반응 로그를 중심으로, 이를 설명하는 캠페인 속성, 고객 속성, 그리고 외부 시점 정보가 포함된 4개의 관계형 테이블로 구성된다.

문장유사율: 0%

그리고 캠페인 유형은 고객의 프로파일이나 프로모션과 직접적인 연관이 있는 것들로 제한하였고, 단순한 거래 진행현황과 같은 정보들의 캠페인은 제외하였다.

문장유사율: 0%

이들은 전처리 과정을 통해 통합되어, 후술할 도메인 지식기반 파생변수(Feature Engineering) 산출의 기초가 된다.

문장유사율: 0%

마케팅 구매 여부 예측에 활용되는데이터테이블 변수 구성 및 정의의 데이터의 이해도를 높이기 위해, 본 연구의 모델링에 활용되는데이터의 구조적 특징은 크게 세 가지로 요약된다.

문장유사율: 0%

첫째, 단순히 특정 캠페인에 대한 반응이 포함된 것이 아니라 다양한 캠페인 반응을 포함한 고객의 다중 접점(Multi-touch) 여정을 반영하고 있다.

문장유사율: 0%

고객은 단 한번의 메시지로 즉시 구매하지 않는다. 이메일, 앱 푸시, SMS 등 시간의 흐름에 따라 다양한 채널을 통해 메시지가 누적되며, 고객이 최종적으로 설득되거나 이탈하는 과정이 데이터에 포함되어 있다.

문장유사율: 0%

둘째, 최종적인 목표인 구매 여부 클래스의 불균형이 극심하다. 전체 발송 메시지 중 실제 구매로 이어진 비율은 약 0.12%에 불과하기 때문이다.

문장유사율: 0%

마지막으로 셋째는 메시지, 캠페인, 고객, 시간 정보가 서로 다른 식별자(Key)로 분리되어 있다.

문장유사율: 0%

따라서 이를 통합하여 개별 고객이 시간 흐름에 따라 겪는 경험을 하나의 시퀀스 형태로 재구성하는 전처리 과정이 필수적이다.

문장유사율: 0%

이러한 과정은 최종적으로 병합되어 향후 기술할 지식기반 파생변수 생성의 기초가 된다.

문장유사율: 0%

비록 본 데이터가 특정 지역의 플랫폼에서 수집되었으나, 해당 플랫폼은 다수의 상품 카테고리들을 포괄하는 종합 이커머스 형태를 띠고 있다.

문장유사율: 0%

또한, 고객 행동 로그가 전형적인 메시지 수신 -> 오픈 -> 클릭 -> 구매로 이어지는 깔대기 구조로 구성되어 있고 마케팅에 대한 반응 곡선도 글로벌 유사한 롱테일(Long-tail) 분포를 따르고 있기에, 본 연구의 결과가 일반적인 이커머스 환경에 충분히 일반화될 수 있다

문장유사율: 0%

고 판단된다 2. 도메인 지식기반 파생변수 프레임워크 본 연구는 제2장에서 고찰된 마케팅이론의 추상적 개념들을 머신러닝 모델이 직접 학습 가능한 정량적 입력 벡터로 변환하는 이론의 조작화에 중점을 두었다.

문장유사율: 0%

고객의 내재적 심리와 행동 메커니즘을 데이터로 포착하기 위해, 본 연구는 (1) 구매 및 재구매 주기 패턴, (2) 시계열적 역동성 및 메시지 반응 패턴, (3) 마케팅 피로도 및 상호작용 강도 패턴, (4) 마케팅 콘텐츠 신규성 및 경로 효율성 패턴의 4가지 차원으로 구성된 파생변수 프레임워크를 제안한다.

문장유사율: 0%

1) 구매 및 재구매 주기 패턴 본 변수군은 고객 생애가치 및 RFM 이론을 기반으로, 고객의 구매 주기가 갖는 시계열적 규칙성을 정량화 하기 위해 설계되었다.

문장유사율: 0%

첫째, 재구매 준비도를 통해 고객별 구매 간격의 이질성을 반영하였다. 소비자는 저마다 고유한 소비 주기를 가진다.

문장유사율: 0%

이를 모델링 하기 위해 개별 고객의 과거 구매 간격 분포를 정규분포()로 가정하고, 현재 경과 시간()이 고객의 평균 주기()와 얼마나 일치하는지를 나타내는 확률 밀도 유사도를 산출하였다.

문장유사율: 0%

엄밀한 통계적 모델링에서는 구매 간격을 감마 분포 등으로 가정하나, 본 연구에서는 머신러닝 모델의 피처로서의 직관성과 연산 효율성을 고려하여 중심 극한 정리에 근거한 정규분포 근사를 채택하였다.

문장유사율: 0%

따라서, 이 값이 1에 근접할수록, 단순히 시간이 많이 경과한 것이 아니라, 고객의 내재적 습관에 기반하여 평소 습관대로라면 구매 임박 시점에도 달했음을 의미하는 강력한 신호로 해석된다.

문장유사율: 0%

고객 고유의 구매 사이클에 맞춰 적절한 타이밍에 메시지가 도달했음을 의미한다. 둘째, 구매후 불응기(Post-purchaseRefractory Period)를 정의하여 소비 직후의 효용 만족도 효과를 반영하였다.

문장유사율: 0%

구매 직후 일정기간 동안은 재구매 확률이 급격히 감소하는 재고 효과 또는 심리적 냉각기를 반영하기 위해, 경과시간()에 따른 역지수 감쇠 함수(Inverse Exponential Decay)를 적용하였다.

출처 표시 문장

문장유사율: 0%

이메일이나 모바일 기반 마케팅 분야의 실증 연구에 따르면 고객의 반응 대부분은 수 일 내에 발생하며 이후 급격히 감소하는 패턴을 보인다 (Bonfrer and Drèze, 2009; Maileon, 2023). 이러한 패턴을 보수적으로 반영하여 경과 시간을 72시간으로 설정하였다.

문장유사율: 0%

따라서 구매 직후에는 강한 억제 값(0에 수렴)을 갖고 시간이 지날수록 서서히 억제가 풀리는 구조(1에 수렴)를 구현하였다.

문장유사율: 0%

단, 구매이력이 부족하여 개인별 파라미터 추정이 불가능한 신규 고객의 경우, 전체 고객 집단의 글로벌 평균 및 분산을 대체 값으로 적용하여 콜드 스타트 문제를 완화하였다.

문장유사율: 0%

이는 모델이 방금 구매를 완료한 고객에게 불필요한 마케팅 리소스를 투입하는 과적합을 방지하고, 억제기가 해소된 잠재고객을 선별하는 필터링 기제로 작동한다.

문장유사율: 0%

2) 시계열적 역동성 및 메시지 반응 패턴 본 연구는 고객의 반응 확률이 물리적 시간의 흐름에 따라 고정되어 있지 않고 유동적으로 변화한다는 Koren (2009)의 시점 역동성 개념을 확장하여, 고객의 라이프 스타일과 사회 및 경제적 맥락을 반영한 시계열 변수군을 설계하였다.

문장유사율: 0%

이는 "언제 메시지를 보낼 것인가?"에 대한 최적화 문제를 해결하기 위해, 단순한 시각 정보를 고객의 내재적 리듬과 동기화 시키는 과정이다.

문장유사율: 0%

첫째, 원형 통계 기법을 적용하여 고객 고유의 생체 리듬과의 일치도를 수치화 하였다.

문장유사율: 0%

일반적인 선행 시간 척도에서는 23시와 1시 사이의 거리가 멀게 계산되는 불연속성 문제가 발생한다.

문장유사율: 18%

이를 해결하기 위해 본 연구는 24시간을 원형 좌표계(Circular Coordinate System)로 매핑하고, 고객의 과거 반응(Open, Click, Purchase) 시점을 매핑하여 고객별 평균 선호 시간 벡터()를 산출하였다.

문장유사율: 0%

평균 선호 시간 산출시, 시간의 연속성을 보존하기 위해 각 시점을 단위 벡터로 변환하여 벡터 평균을 산출하였다.

KCI 논문 | 제목 : PageRank 특징을 ... | 저자 : 윤지영(가... | 발행년 : 2021.08

이를 해결하기 위해 본 연구는 페이지랭크(PageRank) 특징을 활용한 RDP기반 내부전파경로 탐지 및 SHAP를 이용한 설명가능한 시스템을 제안한다.

문장유사율: 0%

현재 발송 시점()과 고객 선호 시간() 간의 편차는 코사인 거리를 활용하여 다음과 같이 0~1 범위로 정규화된다.

문장유사율: 0%

이 지표가 0에 수렴할수록, 현재 메시지가 고객이 가장 활발하게 반응하는 주 활동시간대에 정확히 도달했음을 의미한다.

문장유사율: 0%

둘째, 경제적 캘린더 효과를 반영하기 위해 소득 가용 시점(Payday)까지의 근접도를 모델링 하였다.

문장유사율: 0%

고객의 소비심리는 가처분소득이 일시적으로 증가하는 급여일이나 월말 정산기 전후로 고조되는 소득 효과의 영향을 받는다. 이를 반영하기 위해 특정 경제 이벤트시점()을 기준으로, 시간의 근접성에 따라 가중치가 비선형적으로 변화하는가우시안 범프 함수(Gaussian Bump Function)를 적용하였다.

문장유사율: 0%

이는 급여일 당일에 가까워질수록 값이 1로 급증하고 멀어지면 0으로 수렴하는 구조로, 특정 시기에 폭발적으로 증가하는 잠재 구매력을 모델이 학습할 수 있도록 유도하는 확률적 장치이다.

문장유사율: 0%

셋째, 거시적인 사회와 문화적 시점 변수를 추가하여 요일 및 계절성 효과를 통제하였다.

문장유사율: 0%

주말과 평일의 행동양식 차이, 그리고 월초의 탐색적 소비와 월말의 목적형 소비패턴의 차이를 반영하기 위해 주차정보 및 분기말 잔여일수를 변수화하였다.

문장유사율: 0%

결론적으로 이 변수군은 단순한 텍스트 정보를 넘어, 메시지가 도달한 순간의 시공간적, 경제적 맥락을 입체적으로 재구성 함으로써 모델의 예측 정확도를 보정하는 역할을 수행한다.

문장유사율: 0%

3) 마케팅 피로도 및 상호작용 강도 패턴 본 변수군은 고객의 인지적 자원 한계와 자극에 대한 습관화 이론을 기반으로, 마케팅 노출 빈도와 고객 반응 사이의 비선형적 관계를 모델링한다.

문장유사율: 0%

이는 본 연구의 실험에서 예측 성능에 가장 높은 기여도를 보인 핵심 요인으로, 고객의 즉각적 반응 성향과 이탈 위험을 동적으로 포착하도록 설계되었다.

문장유사율: 0%

첫째, 채널별 침해성을 고려하여 마케팅 누적 피로도 함수를 제안하였다. 메시지 수신은 고객의 주의를 환기시키지만, 임계치를 초과한 빈번한 노출은 심리적 저항을 유발한다.

문장유사율: 0%

본 연구는 이러한 기억의 망각 및 피로 해소 과정을 지수 감쇠 모델로 수식화하였다. 특정 시점()에서의 누적 피로도()는 과거 수신된 모든 메시지의 잔존 영향력의 합으로 정의된다.

문장유사율: 0%

메시지 수신 직후에는 피로도가 급증하지만, 시간이 지나면서서히 해소되는 기억의 망각 과정을 수식으로 구현한 것이다.

문장유사율: 0%

이때 피로도가 사라지는 속도인 는 채널()의 피로도 반감기를 의미하며, 매체의 침해성과 고객 반응 임계점에 대한 선행 연구에 근거하여 차등 적용하였다.

문장유사율: 0%

둘째, 행동의 최신성과 휴면 상태를 이원화하여 분석하였다. 피로도 변수가과다 노출의 부작용을 측정한다면, 최신성 변수는 접점의 부재 기간을 측정한다.

문장유사율: 0%

각 채널별 마지막 수신 시점으로부터의 경과시간()을 산출하여, 이 값이 임계치를 초과할 경우 고객이 비활성 상태에 진입했음을 시사한다.

문장유사율: 0%

반면, 장기 휴면 고객에게 발송되는 메시지는 피로도가 아닌 재인식 효과를 유발할 수 있으므로, 모델은 피로도와 최신성 변수의 상호작용을 통해 이를 구분하여 학습하게 된다.

문장유사율: 0%

셋째, 슬라이딩 윈도우(Sliding Window) 기법을 적용하여 단기 상호작용 밀도를 산출하였다.

문장유사율: 0%

발송 시점 기준 직전 30일간의 이동구간 내에서 발생한 메시지 열람 및 클릭의 빈도와 비율(CTR)을 집계하였다.

문장유사율: 0%

이는 고객의 관심사가 일시적으로 고조된(High-intent) 상태인지, 아니면 지속적으로 반응하지 않는(Low-intent) 상태인지를 정량화하여 단기 구매 전환 확률을 추정하는 핵심 지표로 활용된다.

문장유사율: 0%

넷째, 데이터 희소성 문제를 완화하기 위해 전역적 반응 통계를 보완재로 주입하였다.

문장유사율: 0%

신규 고객(Cold Start)이나 이력이 부족한 롱테일 고객의 경우, 개별 행동 로그만으로는 예측의 불확실성이 높다. 이를 보정하기 위해, 현재 발송 예정인 메시지 주제와 채널 조합에 대해 전체 고객 집단이 최근 30일간 보인 평균 반응률을 변수로 활용하였다.

문장유사율: 0%

이는 개별 정보가 부재할 때 시장의 최신 트렌드 정보를 통해 예측의 공백을 메우는 베이저안 사전 확률(Bayesian Prior)과 유사한 역할을 수행한다.

문장유사율: 0%

4) 마케팅 콘텐츠 신규성 및 경로 효율성 패턴 본 변수군은 메시지의 내용적, 구조적 적합성을 평가하여, 이미 형성된 고객의 구매 의도를 실제 전환으로 유도하는 설득 효율성을 최적화하기 위해 설계되었다.

문장유사율: 0%

첫째, 콘텐츠의 주제 신선도를 정량화하여 정보의 한계 효용 효과를 반영하였다. 동일한 마케팅 메시지의 빈번한 반복 노출은 고객의 주의력을 저하시키고 광고 맹목 현상을 유발한다.

문장유사율: 0%

이를 방지하기 위해, 특정 주제()에 대한 최근 7일간의 노출 빈도()와 마지막 노출 후 경과시간()을 결합한 복합 감쇠 함수를 적용하였다.

문장유사율: 0%

이때, 해당함수의 감쇠 속도를 결정하는 핵심 하이퍼 파라미터인 피로 강도()와 반감기()는 앞서 논의한 채널별 침해성 이론에 근거하여 선행연구의 방식을 활용하였다.

문장유사율: 0%

이 점수()가 낮을수록 해당 주제가 과포화 상태를 의미하며, 반대로 높을수록 고객에게 새로운 자극으로 인지될 가능성이 높음을 시사한다.

문장유사율: 0%

이는 마케터가 고객의 피로도를 유발하지 않는 범위내에서 콘텐츠의 다양성을 제어하는 지표로 활용된다.

문장유사율: 0%

둘째, 성공 경로 일치도를 산출하여 전환 가능성을 구조적으로 예측하였다. 이는 현재 고객이 경험하고 있는 채널 조합 접점이 과거 구매자들의 전형적인 성공 경로와 얼마나 구조적으로 유사한지를 측정하는 것이다.

문장유사율: 0%

본 연구는 실제 구매가 발생하기 직전 5단계의 채널 흐름을 전환 프로토타입()으로 정의하고, 현재 고객의 최근 경로()와의 유사도를 자카드 계수(Jaccard Coefficient)를 통해 산출하였다.

문장유사율: 0%

이 지표는 단순한 개별 채널의 빈도 분석을 넘어, 구매를 유발하는 검증된 설득의 구조를 현재 고객이 따르고 있는지를 판단하는 패턴 매칭 변수이다.

문장유사율: 0%

셋째, 캠페인 속성 재현도를 통해 과거 성공 요인의 재현 여부를 평가하였다. 고객이 마지막으로 구매를 결정했을 당시 반응했던 캠페인의 속성 벡터(할인율, 이모지 사용, 톤 앤 매너 등)를 로 정의하고, 현재 발송 예정인 메시지의 속성 벡터()와의 이진 매칭 점수를 계산하였다.

문장유사율: 0%

이 값이 1에 근접할수록 현재 메시지가 해당 고객의 과거 구매 트리거를 정확히 재현하고 있음을 의미하며, 이는 개인화된 리타겟팅의 정밀도를 높이는 결정적인 신호로 작용한다.

문장유사율: 0%

단, 구매력이 없는 고객의 경우, 해당 변수 값을 임의의 통계치를 반영하지 않고 0으로 할당하여 특성의 왜곡을 방지하였다.

문장유사율: 0%

3. 데이터 전처리 및 데이터셋 구축 본 연구는 원시 로그 데이터가 내포한 고차원성과 희소성, 그리고 클래스 불균형(Class Imbalance) 문제를 해결하고, van Tol (2024)이 제안한 베이스라인 모델의 재현성을 확보하기 위해 엄격한 전처리 파이프라인을 구축하였다.

문장유사율: 0%

분석 데이터셋은 노이즈 제거 및 학습 효율성 최적화를 위해 다음 5단계의 정제 과정을 거쳐 생성되었다.

문장유사율: 0%

1) 변수 표준화 및 특징 선택 비정형 로그 데이터의 시계열적 일관성을 확보하기 위해, 모든 시간 변수를 UTC 기준의 표준 포맷으로 통일하였다.

문장유사율: 0%

이어지는 변수 선택 단계에서는 모델의 강건성(Robustness)을 저해하는 요인을 배제하기 위해 다음 두 가지 기준을 적용하였다.

문장유사율: 0%

첫째, 무정보성 변수의 제거이다. 전체 관측치의 99% 이상이 결측이거나 단일 값(Constant)을 갖는 변수는 종속변수에 대한 판별력이 부재하므로 분석 대상에서 제외 하였다.

문장유사율: 0%

둘째, 정보 중복성 및 사후 변수의 배제이다. 제목_길이와 같은 단순 메타 텍스트 변수는 모델 복잡도만 증가시킬 뿐 고객 여정의 본질적 정보를 제공하지 않는다는 선행 연구의 논의에 따라 제거하였다.

문장유사율: 0%

또한, 수신_불가 여부와 같이 발송 이후에 결정되는 기술적 로그 변수는 사전 예측 시점에 가용하지 않은 정보이므로 입력 변수로서의 타당성이 결여되어 제외하였다.

문장유사율: 0%

2) 빈도 기반 화이트 리스트를 통한차원 축소 범주형 변수의 과도한 카디널리티(Cardinality) 증가는 데이터 희소성을 심화시켜 차원의 저주(Curse of Dimensionality)를 유발한다.

문장유사율: 0%

이를 완화하기 위해 본 연구는 빈도 기반의 화이트리스트(Whitelist) 접근법을 적용하였다.

문장유사율: 17%

전체 범주를 원-핫 인코딩(One-Hot Encoding)하는 대신, 상위 빈도를 점유하는 주요 범주(이메일 도메인 상위 3개, 접속 플랫폼상위 5개 기기)만을 독립 변수로 보존하고, 나머지 롱테일(Long-tail) 범주는 기타(Other)로 통합하였다.

KCI 논문 | 제목 : 딥러닝 알고리즘을... | 저자 : 김서연(고... | 발행년 : 2022.04

이외에도 범주형 변수를 원-핫 인코딩 (one-hot encoding)으로 변환하였고, 전체 데이터에 표준 스케일 리(standardscaler)를 적용하여 분석을 위한 전처리과정을 수행하였다.

KCI 논문 | 제목 : 고객의 속성별 중... | 저자 : 우수현(경... | 발행년 : 2025.03

여기서, d는 형성된 밀집 벡터의 차원수를 의미하며, m과 n은 고유한 고객 및 레스토랑의 수를 나타낸다. 그리고 와 는 고객과 레스토랑 상호작용을 나타내는 원-핫 인코딩(One-hot Encoding) 벡터를 의미한다.

KCI 논문 | 제목 : 온라인 학습 데이... | 저자 : 권연하(건... | 발행년 : 2025.03

이와 같이 계산된 값은 원-핫 인코딩(one-hot encoding)을 거쳐 벡터로 변환되며, 예를 들어 상호작용 값(값값)이 8인 경우에는 (0,0,0,0,0,0,0,1,0,0), 3인 경우에는 (0,0,1,0,0,0,0,0,0,0)으로 표현된다.

문장유사율: 0%

이는 변수 공간을 압축하여 모델의 과적합을 방지하고 일반화 성능을 높이기 위한 조치이다.

문장유사율: 0%

3) 고객 여정 시퀀스 통합 및 비정규화 관계형 데이터베이스 구조로 파편화된 테이블을 고객 ID와 캠페인 ID를 키(Key)로 병합(Left Join)하여, 단일 사용자의 시계열적 행동 흐름을 추적 가능한 시퀀스 데이터로 비정규화하였다.

문장유사율: 0%

특히 외부 변수인 공휴일 정보와의 병합 과정에서는 구매 여부에 따른 데이터 구조적 정합성을 유지하기 위해, 매핑 테이블을 선제적으로 구조화 함으로써 병합 과정에서의 정보 누락 및 불일치 오류를 원천적으로 차단 하였다.

문장유사율: 0%

4) 시점 제약을 통한 데이터 누수 방지 본 연구의 예측 시점은 메시지 발송 시점()이다. 따라서 모델의 입력은 반드시 발송 직전 시점()까지 관측된 과거 데이터로만 한정되어야 한다.

문장유사율: 0%

미래의 정보가 학습에 포함되는 데이터 누수(Data Leakage) 혹은 선견 편향(Look-ahead Bias)을 방지하기 위해, 본 연구는 엄격한 시점 제약을 적용하였다.

문장유사율: 0%

이에 따라 누적이력, 최신성, 캠페인 맥락 변수는 모두 발송 직전까지의이력만으로 산출 되었으며, 사후적으로 확정되는 정보는 학습에서 철저히 배제되었다.

문장유사율: 0%

5) 하이브리드 샘플링 및 데이터 분할 원본 데이터는 구매 전환율이 약 0.12%에 불과한 극심한 클래스 불균형을 보인다. 이는 모델이다수 클래스(Non-purchase)로 편향되는 주 원인이 되므로, 본 연구는 van Tol (2024)의 표본 주입(Sample Infusion)에 기반한 하이브리드 샘플링 전략을 채택하였다.

문장유사율: 0%

우선, 소수 클래스의 정보 손실을 방지하기 위해 1.7억 건 중 실제 구매(Positive) 로그는 전량 보존하였다.

문장유사율: 0%

반면, 비구매는 Thabtah et al. (2020)이 제안한 불균형 학습의 최적 비율을 고려하여, 데이터셋 내 양성 비율이 약 10%에 수렴하도록 무작위 언더샘플링을 수행하였다.

문장유사율: 0%

다음으로, 최종 구축된 10만 건의 데이터셋은 사용자 집단 간의 편향을 제거하기 위해 무작위 서플링 후 7:1.5:1.5 (Train/Validation/Test) 비율로 분할 하였다.

문장유사율: 0%

본 연구는 모델 간의 상대적 우위를 비교하는데 목적이 있으므로 평가 데이터의 클래스 비율도 통제하였다.

문장유사율: 0%

따라서 실무 적용 시에는 보정이 필요할 수 있다. 최종적으로 분할된 데이터는 모델 학습, 하이퍼 파라미터 최적화, 그리고 일반화 성능의 객관적 검증을 위한 표준 절차를 따른 것이다.

문장유사율: 0%

4. 예측 알고리즘 아키텍처 및 실험 설계 본 연구의 분석 대상인 고객 여정 로그는 전환(Conversion) 의사결정이 단일 변수에 의해 선형적으로 설명되지 않고, 채널 간의 상호작용, 시간의존성, 그리고 특정 시점에 국소적으로 발현되는 반응 패턴에 의해 복합적으로 결정된다는 특징을 갖는다.

문장유사율: 0%

그러나 기존 연구들은 정형 요약 변수(Summary Statistics)에 의존한 트리모델링에 치중하거나, 시퀀스 모델을 적용하더라도 입력 단계에서 임베딩이나 차원 축소 기법을 과도하게 적용함으로써, 원시 로그가 내재한 정보량의 한계를 구조적으로 은폐하는 경향이 있었다.

문장유사율: 0%

이에 본 연구는 도메인 지식기반의 피쳐 엔지니어링이 시계열적 맥락을 충분히 내재화했을 때, 전통적 앙상블 모델과 딥러닝 아키텍처가 각각 어떠한 기제로 예측 성능을 발현하는가를 규명하고자 한다.

문장유사율: 0%

이를 위해 총 8종의 알고리즘을 동일한 데이터 맥락 하에서 구축하고, 각 모델의 최적화를 위한 주요 하이퍼 파라미터를 베이지안 최적화(Bayesian Optimization)를 기반으로 튜닝하고 교차 검증을 통해 성능 우위의 원천을 실증적으로 비교 분석하였다.

문장유사율: 0%

1) 트리 기반 앙상블 알고리즘: Random Forest, XGBoost 딥러닝 모델의 효율성을 객관적으로 평가하기 위한 기준점(Benchmark)으로서, 정형 데이터 분석에서 강력한 일반화 성능이 입증된 트리 기반 앙상블 모델을 채택하였다.

문장유사율: 0%

Random Forest (RF)는 부트스트랩 표본추출을 통해 다수의 결정 트리를 생성하고 그 결과를 결합함으로써, 개별 모델의 편향을 유지하면서 분산(Variance)을 효과적으로 감소시키는 전략을 취한다.

문장유사율: 0%

반면, XGBoost (XGB)는 그래디언트 부스팅 프레임워크를 기반으로 이전 트리의 잔차(Residual)를 순차적으로 보정하며, 희소 데이터 처리에 특화된 알고리즘 구조와 병렬 연산 최적화를 통해 실무적 효율성을 제공한다.

문장유사율: 0%

이들 트리 계열 모델은 입력 데이터의 순차적 흐름을 직접 학습하지는 않는다. 그러나 본 연구에서 설계한 도메인 지식기반 피쳐(구매 주기 패턴, 피로도 등)가 이미 시간적 구조를 정형 신호로 요약하고 있으므로, 트리 모델이 변수 간의 교차 상호작용과 비선형 경계를 학습함으로써 어느 수준까지 시계열 예측 성능을 근사할 수 있는지 검증하는 것은 중요하다.

문장유사율: 0%

이는 향후 딥러닝 모델이 제공할 시퀀스 학습의 이득을 식별하는 핵심 비교 기준이 된다.

문장유사율: 0%

2) 단일 딥러닝 알고리즘: MLP, CNN, RNN, LSTM 본 연구는 정형화된 도메인 변수들이 신경망에 입력되는 텐서 구조에 따라 어떠한 예측 특성을 보이는지 규명하기 위해, 데이터의 위상학적 표현 방식을 달리한 네 가지 단일 아키텍처를 구축하였다.

문장유사율: 0%

첫째로, Multi-Layer Perceptron (MLP)는 가장 기본적인 전방향 신경망(Feed-forward NeuralNetwork)으로, 입력 데이터를 () 형태의 2차원 매트릭스로 처리한다.

문장유사율: 0%

이는 모든 입력 변수가 은닉층의 뉴런과 완전히 연결되는 구조로, 변수의 배열 순서나 국소적 관계보다는 변수 전체의 전역적 상호작용과 비선형 결합 강도가 예측력을 좌우하는 상황에서 베이스라인(Baseline) 역할을 수행한다.

문장유사율: 0%

둘째로, Convolutional NeuralNetwork (CNN)은 도메인 변수들 간의 국소적 관계를 포착하기 위해 입력을 ()의 3차원 텐서로 재구성하고 1D-CNN을 적용하였다.

문장유사율: 0%

커널(Kernel)이 피쳐 축을 따라 슬라이딩하며 연산을 수행함으로써, 전체 변수의 전역적 결합보다는 인접한 변수 그룹 내에 존재하는 국소적 패턴이나 미세 변동이 전한 예측의 핵심 신호로 작용할 때 유효한 접근법이다.

문장유사율: 0%

셋째로, RNN은 순환구조를 통해 은닉 상태(Hidden State)를 갱신하지만, 본 연구에서는 정형 데이터의 속성을 고려하여 입력 차원을 ()로 설정하였다.

문장유사율: 0%

이는 전체 피쳐 집합을 단일 시점(Time-step=1)의 벡터로 간주하여 처리하는 방식으로, 시계열적 필립 없이 RNN 셀 고유의 비선형 활성화 기제가 MLP 대비 어떠한 표현 학습의 차이를 만드는지 검증하기 위한 설정이다.

문장유사율: 0%

마지막으로 LSTM은 입력, 망각, 출력 게이트로 구성된 메모리 셀을 통해 정보의 흐름을 제어한다.

문장유사율: 0%

본 연구는 마찬가지로 ()의 입력 구조를 채택하여, 시계열적 장기 의존성 학습보다는 LSTM 셀 내부의 메커니즘이 복잡한 도메인 피쳐(장기성향 vs 단기 피로도) 간의 정보 선택과 소거에 있어 MLP보다 우수한 필터링 능력을 발휘하는지 평가하는데 초점을 둔다.

문장유사율: 0%

3) 하이브리드 딥러닝 알고리즘: CNN-LSTM, RNN-LSTM 단일 아키텍처의 편향을 상호 보완하기 위해, 국소 특징 추출과 시퀀스 모델링을 결합한 하이브리드 구조를 추가로 구성하였다.

문장유사율: 0%

CNN-LSTM에서 CNN을 인코더(Feature Extractor)로, LSTM을 디코더(Sequence Modeler)로 활용하는 구조이다. 입력단에서는 () 구조를 통해 CNN이 피쳐 공간상의 노이즈를 제거하고유의미한 국소 특징을 학습 및 생성한다.

문장유사율: 0%

이후, LSTM은 추출된 특징들의 시퀀스를 입력받아 특징 간의 위상학적 인과관계를 학습한다.

문장유사율: 0%

이는 어떤 반응 변수들이 동쳐서 유의미한 패턴을 만드는지(CNN), 그리고 그 패턴들이 어떤 맥락으로 연결되는지(LSTM)를 계층적으로 분리 학습하는 전략이다.

문장유사율: 0%

다름으로 RNN-LSTM에서는 단순 순환 셀(RNN)과 게이트 기반 셀(LSTM)을 적층(Stacking)한 구조이다.

문장유사율: 0%

하위 계층의 RNN은 입력 벡터의 1차적인 비선형 변환을 담당하고, 상위 계층의 LSTM은 이를 정제하여 최종 의사결정 정보를 생성한다.

문장유사율: 0%

이는 서로 다른 활성화 구조를 가진 셀을 결합함으로써, 단일 모델 대비 더 깊은 표현 공간을 탐색하고 모델의 수용력을 증대시키기 위한 설계이다.

문장유사율: 0%

5. 모델 평가 및 검증 본 연구는 이진 분류(Binary Classification) 문제에서 최적의 구매 예측 모델을 선정하기 위해, 예측의 판별력을 고려한 지표를 적용하였다.

문장유사율: 0%

본 연구의 종속변수인 구매 전환은 전체의 10% 미만을 차지하는 희소 사건이다. 이러한 환경에서 단순 정확도(Accuracy)는 다수 클래스에 편향된 결과를 도출하는 역설을 유발할 수 있다.

문장유사율: 21%

따라서 본 연구는 정확도를 보조지표로 제한하고, 다음의 지표들을 핵심 판단 기준으로 삼았다. 첫째, **정밀도(Precision)**와 **재현율(Recall)**을 통해 마케팅 비용과 기회비용을 평가하였다.

문장유사율: 0%

정밀도는 모델이 구매할 것이라고 예측한 집단 중 실제 구매자의 비율이다. 이는 타겟 마케팅 수행시 마케팅 비용 낭비를 통제하는 척도로, 비용 효율성 관점에서 중요하다.

문장유사율: 0%

재현율은 실제 구매 고객중 모델이 정확히 찾아낸 비율이다. 이는 잠재고객을 놓침으로써 발생하는 기회비용을 최소화하는 척도로, 캠페인의 도달 범위 관점에서의의를 갖는다. 둘째, F1-score를 통해 트레이드오프(Trade-off)를 조정하였다.

문장유사율: 21%

정밀도와 재현율은 상충 관계를 가지므로, **두 지표의 조화 평균인 F1-score**를 산출하여 모델이 불균형 데이터 환경에서도 편향되지 않고 균형 잡힌 예측 성능을 달성했는지 검증하였다.

문장유사율: 0%

셋째, Area Under Curve(AUC)를 통해 모델의 강건성을 검증하였다. 단일 임계값(Threshold)에 의존하는 지표들의 한계를 극복하기 위해, 모든분류 임계값 하에서의 참 긍정률과 거짓 긍정률의 관계를 적분한 AUC 값을 활용하였다.

문장유사율: 0%

이는 클래스 분포가 변화하더라도 모델 고유의 순위결정 능력을 객관적으로 비교할 수 있는 척도이다.

KCI 논문 | 제목 : 혼합물 실험계획법... | 저자 : 권영석(공... | 발행년 : 2024.12

반면, F1-Score는 **정밀도(Precision)**와 **재현율(Recall)**의 조화 평균을 나타내는 지표로, 특히 데이터의 클래스 간 분포가 불균형할 경우에도 모델의 성능을 신뢰성 있게 평가할 수 있다[16]. 특히, 본 연구에서는 혼합물 실험계획법을 활용하여 기저 모델의 가중치 수준별 성능 지수를 구하고, Scheffé 다항회귀 모델을 통해 이를 최적화 하였다.

KCI 논문 | 제목 : 법령 정보 확인 ... | 저자 : 박정원(국... | 발행년 : 2025.06

두 지표의 조화 평균인 F1-Score는 0.665로, 규제 분류에 있어 구분 기준 데이터 생성 데이터N % N % 규제 여부 1. 행정규제 180 46.2 255 65.4 2. 비규제 210 53.8 135 34.6 계 390 100.0 390 100.0 법령 정보 확인 및 규제 판별에서의 생성형 AI 활용 가능성: ChatGPT 기반 탐색적 분석 67 전반적으로 균형 잡힌 성능을 유지하고 있음을 보여준다.

문장유사율: 0%

IV. 실증 분석 결과 1. 지식기반 파생변수의 통계적 특성 및 상관 구조본 절에서는 제안된 도메인 지식기반 파생변수들이 실제 고객 행동 데이터 내에서 어떠한 분포적 특성과 상호의존성을 갖는지 검증한다.

문장유사율: 0%

이는 단순한 기초통계 확인을 넘어, 모델링 입력 변수로서의 판별력과 변수 간 다중 공선성(Multicollinearity) 위험을 사전에 진단하기 위함이다.

문장유사율: 0%

1) 기술통계량을 통한 데이터 특성 진단에 제시된 기초 통계량분석 결과, 본 데이터셋은 이커머스 고객 행동의 전형적인 특징인 희소성과 비대칭성을 뚜렷하게 드러낸다. 주요 발견점은 다음 세가지로 요약된다.

문장유사율: 0%

첫째, 극심한 클래스 불균형과 행동의 양극화이다. 종속변수의 양성 비율은 약 10%에 불과하며(Target Mean=0.10), 마지막 구매 경과일과 마케팅 피로도 등주요 행동 변수들은 강한 롱테일 분포(왜도>2.0)를 보인다. 이는 대다수의 고객이 비활성 상태에 머물러 있는 반면, 소수의 고관여 고객이 전체 반응과 구매를 주도하는 파레토 법칙(Pareto Principle)이 데이터 내에 구조적으로 고착화되어 있음을 시사한다.

문장유사율: 0%

둘째, 잠재적 활성 상태의 확인이다. 재구매 준비도와 구매후 불응기는 평균이 0.95 이상으로 1에 근접한 강한 음의 편포를 보인다. 이는 분석 시점 기준으로 대다수 고객이 물리적 제약을 벗어나, 자극이 주어지면 언제든지 반응할 수 있는 구매 가능 상태에 진입해 있음을 의미한다.

문장유사율: 0%

셋째, 캘린더 효과와 시점의존성이다. 금요일/분기말 근접도와 주말 여부 변수의 분포는 특정 시점에 마케팅 이벤트가 집중되는 운영 패턴을 반영한다.

문장유사율: 0%

이는 고객 반응이 개인의 내재적 욕구뿐만 아니라, 기업의 시점 마케팅 전략에 의해 외생적으로 조절되고 있음을 통계적으로 뒷받침한다.

문장유사율: 0%

도메인 지식기반 파생변수의 기초통계량 요약 2) 변수 간 상관 구조 및 모델링 시사점 주요 파생변수 간 피어슨 상관관계 히트맵의 피어슨 상관관계 히트맵(Heatmap)을 통해 도출된 구조적 특징은 다음과 같다. 첫째로 타겟 변수와의 약한 선형 상관성이 관찰된다.

문장유사율: 0%

대부분의 파생변수 간 선형 상관계수는 0.20 보다 작은 수준의 미약한 관계를 보였다. 유일하게 구매후 불응기만이 유의미한 음의 상관(-0.37)을 보여 행동경제학적 냉각기 효과를 입증하였다.

문장유사율: 0%

이러한 낮은 선형 상관성은 구매 의사결정이 단일 요인의 선형적 합이 아니라, 피로도, 시점, 콘텐츠 등 다차원 요인의 복잡한 비선형적 상호작용에 의해 결정됨을 강력히 시사한다.

문장유사율: 0%

이는 본 연구가 선형 회귀가 아닌, 고차 상호작용을 포착할 수 있는 트리 기반 앙상블이나 딥러닝 모델을 채택해야 하는 당위성을 제공한다.

문장유사율: 0%

둘째로, 각 카테고리 내 변수들은 중간 수준의 상관($0.4 < r < 0.7$)을 보였으나, 서로 다른 카테고리 간 상관계수는 대부분 0.2 미만으로 낮게 나타났다.

문장유사율: 0%

이는 제안된 4가지지식 축이 서로 중복되지 않는 독립적인 정보를 제공하고 있음을 의미하며, 다중공선성 위험 없이 모델의 설명력을 극대화할 수 있는 구조적 타당성을 확보하였음을 확인하였다.

문장유사율: 0%

일부 변수인 불응기와 재구매 준비도는 0.73으로 다소 상관계수가 높지만 연구에서 사용하는 머신러닝 및 딥러닝 알고리즘은 다중공선성에 상대적으로 강건(Robust)하기 때문에 별도로 제거하지 않고 학습에 반영하였다.

문장유사율: 0%

2. 실험 설계 및 분석 프레임워크 본 장에서는 제1장에서 제기된 연구 질문(RQ1~RQ3)을 체계적으로 검증하기 위한 실험 절차를 기술한다.

출처 표시 문장	문장유사율: 0%
분석 프레임워크는 크게 두 단계로 구성된다. 첫째, 선행연구(van Tol, 2024)의 실험 환경을 복원하여 베이스라인 모델(RNN-LSTM)의 성능을 재현하고, 원시 로그 기반의 알고리즘별 기초 성능을 비교한다(RQ3).	
	문장유사율: 0%
둘째, 2장에서 제안된 4가지 도메인 지식 피쳐 세트를 단계적으로 투입하여 각 지식 축의 한계 기여도(Marginal Contribution)와 최적의 피쳐 조합을 규명한다(RQ1, RQ2).	
	문장유사율: 0%
이를 위해 3.3절에서 구축된 약 10만 건의 학습 데이터셋()을 활용하였으며, 종속변수는 메시지 발송 후 24시간 이내 구매 전환 여부이다.	
	문장유사율: 0%
비교 대상 알고리즘은 모델의 구조적 편향을 배제하기 위해, 전통적 머신러닝(RF, XGB)부터 딥러닝(MLP, RNN, LSTM, CNN), 하이브리드 모델(CNN-LSTM, RNN-LSTM)까지 총 8종으로 구성하였다.	
	문장유사율: 0%
제안하는 방법론의 프레임워크 1) 하이퍼 파라미터 최적화 및 검증 프로토콜 본 연구는 모든 비교 모델에 대해 동일한 학습, 검증 프로토콜을 적용하여 실험의 공정성을 확보하였다.	
	문장유사율: 21%
데이터 분할은 3.3절에서 정의한 시계열적 순서를 보존한 학습(Train), 검증(Validation), 테스트(Test) 분할을 유지하되, 학습 단계에서는 특정 데이터 분포에 의한 과적합을 방지하기 위해 5-fold 교차검증을 적용하였다.	KCI 논문 제목 : 2차원 신체 관절... 저자 : 박재욱(국... 발행년 : 2024.12 이를 위해, 데이터셋의 라벨을 활용하여 정상 동작과 오류 동작을 구분하였으며, 학습(Train), 검증(Validation), 테스트(Test) 세트로 분할하였다.
	문장유사율: 0%
이는 특정 데이터 조합에 의해 우연히 성능이 높게 나타나거나 미래의 정보가 과거 학습에 반영되는 편향을 원천 차단하기 위함이다.	
	문장유사율: 0%
각 알고리즘의 최적 하이퍼 파라미터 탐색을 위해, 기존의 무작위 탐색 대비 수렴 속도와 탐색 효율이 우수한 Optuna 기반의 베이지안 최적화(Bayesian Optimization) 기법을 도입하였다.	
	문장유사율: 0%
탐색 공간(Search Space)은 각 모델의 구조적 특성을 반영하여 개별 설정되었으며(), 검증 세트의 AUC를 최대화하는 파라미터 조합을 최적해로 선정하였다.	
	문장유사율: 0%
아울러 Early Stopping과 Pruning 기법을 병행하여 불필요한 연산 비용을 절감하고 일반화 성능을 극대화하였다.	
	문장유사율: 0%
모델링의 하이퍼 파라미터 탐색범위 및 최적값 요약 2) 선행연구 베이스라인 복원 및 성능 정합성 검증 제안하는 도메인 지식 피쳐의 효과를 엄밀하게 측정하기 위해서는, 비교의 기준점이 되는 베이스라인 성능의 신뢰성이 선행되어야 한다.	
	문장유사율: 0%
이에 van Tol (2024)의 실험 설계를 충실히 복원하여, 본 연구의 베이스라인 모델 성능이 선행연구와 정합성을 갖는지 검증하였다.	
	문장유사율: 0%
비교 분석 결과, 공통 모델인 RNN-LSTM의 성능은 선행연구(F1: 0.7023, AUC: 0.8530)와 본 연구(F1: 0.6821, AUC: 0.9630) 간에 유의미한 구조적 유사성을 보였다.	
	문장유사율: 0%
F1-score는 절대적 수치에서 소폭의 차이가 관찰되었으나, 이는 클래스 불균형 비율 및 샘플링 방식의 미세한 차이에 기인한 허용 범위내 변동으로 판단된다.	
	문장유사율: 0%
두 결과 모두 불균형 데이터 환경에서 0.68~0.70 수준의 준수한 조화 평균 성능을 달성하였다.	
	문장유사율: 0%
AUC는 선행연구 대비 높게 측정되었는데, 본 연구가 적용한 정교한 전처리와 최적화 과정이 모델의 순위결정 능력을 효과적으로 개선했음을 시사할 수 있다.	

문장유사율: 0%

또한, 선행연구 대비 본 연구는 Specificity가 높고 Recall이 다소 낮은 경향을 보였으나, 이는 분류 임계값 설정의 차이에 따른 자연스러운 현상이며 모델의 본질적 예측력 이상 이합을 의미하지 않는다.

문장유사율: 0%

나아가 본 연구는 RNN-LSTM 외 7종 알고리즘에 대해서도 베이스라인 성능을 측정하였다().

문장유사율: 0%

분석 결과, 도메인 지식이 부재한 원시 로그 환경에서는 복잡한 딥러닝 모델이 전통적인 트리 기반 앙상블 모델 대비 뚜렷한 성능 우위를 보이지 못하거나, 오히려 소폭 낮은 성능을 기록하였다.

문장유사율: 0%

이는 정제되지 않은 시퀀스 데이터 만으로는 딥러닝의 표현 학습능력이 충분히 발휘되지 못함을 실증하는 결과로, 본 연구가 제안하는 도메인 지식기반 피쳐 엔지니어링의 필요성을 역설적으로 뒷받침한다.

문장유사율: 0%

종합적으로, 본 연구의 실험 환경은 선행연구의 성능 분포를 충실히 재현하고 있으며, 이후 수행될 도메인 지식 확장의 효과를 검증하기 위한 신뢰할 수 있는 기준으로 기능함을 확인하였다.

문장유사율: 0%

알고리즘별 베이스라인 적용성능 비교 3. 지식기반 변수의 한계 기여도 분석 본 절에서는 3.2절에서 제안한 4가지 도메인 지식 축이 예측 성능에 미치는 고유한 기여분을 검증하기 위해, 베이스라인 모델에 각 피쳐 세트를 개별적으로 추가하는 절삭 실험을 수행하였다.

문장유사율: 0%

이는 각 지식 요인이고객 행동의 어떤 차원을 설명하며, 모델의 판별 구조를 어떻게 재구성하는지 정량적으로 규명하기 위함이다.

문장유사율: 0%

1) 구매 및 재구매 주기 패턴 구매 및 재구매 주기 패턴 변수군은 고객의 평균 구매 간격과 재구매 준비도와 같이, 개별 고객이 지니는 장기적 기저 확률(Long-term Base Probability)을 요약하는 지표이다.

문장유사율: 0%

이 변수군은 이론적으로 고객별 기본 구매 확률을 정의해 줄 수 있지만, 단기전환 예측에서 어느 정도의 실질적 변별력을 제공하는지는 경험적으로 확인할 필요가 있다.

문장유사율: 0%

은 구매 및 재구매 주기 변수를 포함하기 전후의 모델 성능을 비교하여 F1-score 기준으로 내림차순 정리한 결과를 제시한다.

문장유사율: 0%

머신러닝 알고리즘이 상위에 위치하고 딥러닝 알고리즘들은 뒤따르고 있다. 또한 Accuracy와 AUC 결과에서도 마찬가지로의 순위를 차지하고 있다.

문장유사율: 0%

그리고 일부 성능이 낮은 딥러닝 알고리즘을 제외한 모든 알고리즘에서 구매 및 재구매 주기 패턴 변수를 포함한 이후 전반적인 분류 성능 지표가 일관되게 개선되는 양상이 확인되었다(Bold). 평균적으로 Precision과 Recall은 각각 0.64%와 10.39%의 상승이 있었으며, 이 둘의 조화평균인 F1-score는 6.34%로 대폭 향상되었다.

문장유사율: 0%

유사하게 Accuracy와 AUC도 1% 미만의 상승을 보인다. 이는 모델이 단순히 최근 클릭이력에 의존하던 근시안적 시각에서 벗어나, 고객의 내재적 소비 주기를 학습함으로써 아직 클릭하지 않았으나 구매 임계점에 도달한 잠재고객을 식별하는 능력이 강화되었음을 의미한다.

문장유사율: 0%

특히 비시계열 모델인 RF/XGB에서도 뚜렷한 성능 향상이 관찰된 것은, 주기성 정보가 시간의 흐름을 정형화된 신호로 효과적으로 요약하여 제공했음을 실증한다.

문장유사율: 0%

종합하면, 구매 및 재구매 주기 패턴 변수는 단순히 개별 시점의 반응을 보완하는 수준을 넘어, 고객 행동의 반복성과 시간적 구조를 체계적으로 반영함으로써 모델의 판별 구조를 재정렬하는 역할을 수행한다.

문장유사율: 0%

구매 및 재구매 주기 패턴 적용에 따른 모델별 성능 2) 시계열적 역동성 및 메시지 반응 패턴 시계열적 반응 패턴 변수군은 발송 시간대, 요일, 급여일이나 분기말 등 발송 시점의 리듬을 정량화한다.

문장유사율: 0%

실무적으로도 언제 보낼 것인가를 정량화한 시점 변수군은 오랜 기간 논의되어온 마케팅 전략 요소로, 모델의 예측 안정성을 보강하는 역할을 할뿐만 아니라 고객이 반응하기 좋은 타이밍을 포착할 경우 동일한 메시지라도 더 높은 성과를 기대할 수 있다.

문장유사율: 0%

는 시계열적 반응 패턴 변수를 기존 모델에 추가했을 때의 성능 변화를 비교하여 F1-score 기준으로 내림차순 정리한 결과를 제시한다.

문장유사율: 0%

마찬가지로 머신러닝 알고리즘이 상위에 위치하고 딥러닝 알고리즘은 뒤따르고 있다. 그리고 CNN 알고리즘을 제외하고 모든 알고리즘에서 시계열적 반응 패턴 변수를 포함한 경우 전반적인 예측 성능이 개선되는 경향이 확인되었다(Bold). 평균적으로 Precision과 Recall은 각각 2.08%와 9.02%의 상승이 있었으며, 이 둘의 조화평균인 F1-score는 5.15%로 대폭 향상되었다.

문장유사율: 0%

유사하게 Accuracy와 AUC도 1% 미만의 상승을 보인다. 이는 구매 이벤트와 같이 상대적으로 희소한 사건을 식별하는데 있어, 반응이력 기반 변수가 분류 민감도를 효과적으로 보완했음을 시사한다.

문장유사율: 0%

특히 XGBoost와 RF와 같은 트리 기반 모델에서는 F1-score가 10% 이상의 두드러진 성능 향상이 나타난 반면, CNN 계열에서는 개선폭이 제한적이었다.

문장유사율: 0%

이는 요일, 급여일 등 명시적 시점 정보가 트리 모델의 분기 조건으로 활용되기 용이한 반면, 국소 패턴을 추출하는 CNN 구조에서는 그 효과가 희석되었기 때문으로 풀이된다.

문장유사율: 0%

결론적으로 시점 변수는 구매 의사결정의 외부적 트리거로서 모델의 민감도를 높이는 기제로 작동한다.

문장유사율: 0%

종합하면, 시계열적 메시지 반응 패턴 변수는 단순히 분류 정확도를 소폭 향상시키는 보조적 요인에 그치지 않고, 모델의 판별 구조 자체를 재구성하는 역할을 수행한다.

문장유사율: 0%

시계열적 메시지 반응 패턴 적용에 따른 모델별 성능 3) 마케팅 피로도 및 상호작용 강도 패턴 마케팅 피로도 패턴 변수군은 고객의 최근 상태를 가장 직접적으로 반영하는 지표로서 가장 드라마틱한 성능 향상을 보인 핵심 변수군들로 구성된다.

문장유사율: 0%

최근 일정기간 동안의 노출 빈도, 채널별 경과시간, 바로 직전 스텝의 반응 여부 등은 고객이 지금이순간 구매할 준비가 되었는지를 판단하는 정보다. 는 마케팅 피로도 패턴 변수를 포함했을 때 각 분류 모델의 성능 변화를 종합적으로 비교하여 F1-score 기준으로 내림차순 정리한 결과를 제시한다.

문장유사율: 0%

본 절에서는 해당 결과를 통해 피로도 기반 특성이 구매 예측성능에 어떠한 구조적 기여를 하는지 분석한다.

문장유사율: 0%

마찬가지로 머신러닝이 상위에 위치하고 있고 딥러닝이 뒤따른다. 그리고 CNN 알고리즘의 AUC를 제외하고 모든 알고리즘과 지표들에서 예측 성능이 상당히 개선되는 경향이 관찰되었다(Bold). 평균적으로 Precision과 Recall은 각각 10.28%와 31.09%의 높은 상승이 있었으며, 이 둘의 조화평균인 F1-score 역시 21.91%로 대폭 향상 되었다.

문장유사율: 0%

유사하게 Accuracy와 AUC도 각각 2.41%와 1.22%의 상대적으로 높은 상승을 보인다. 모든 실험 조건 중 가장 높은 기여도를 기록하였다.

문장유사율: 0%

이러한 결과는 마케팅 피로도 변수가 단순한 보조지표를 넘어, 모델의 판별 구조 전반에 실질적인 영향을 미친다는 점을 시사한다.

문장유사율: 0%

구매 예측에 있어 최신성과 부정적 피로감의 상호작용이 결정적임을 시사한다. 특히 딥러닝 모델뿐만 아니라 모든 알고리즘에서 성능이 비약적으로 상승한 것은, 해당 변수가 지금 반응할 것인가라는 즉각적 전환 의도를 직접적으로 설명하는 강력한 신호임을 입증한다.

문장유사율: 0%

실무적으로이는 과잉 마케팅을 방지하고 적시타겟팅을 가능케 하는 최우선 관리 지표임이 확인되었다.

문장유사율: 0%

종합하면, 마케팅 피로도 변수는 단순한 보조 특성이 아니라, 구매 행동의 누적적·동태적 특성을 구조적으로 반영하는 핵심 설명 변수로 기능한다.

문장유사율: 0%

마케팅 피로도 패턴 적용에 따른 성능 변화 4) 마케팅 콘텐츠 신규성 및 경로 효율성 패턴 마케팅 채널 효과성 및 콘텐츠 패턴 변수군은 현재 발송되는 메시지가 얼마나 새롭고, 과거 성공 조합과 유사하며, 해당 조합이 역사적으로 기록한 반응을 보여왔는지를 정량화한다.

문장유사율: 0%

즉, 이 변수군은 구매 전환을 직접 생성한다기보다, 이미 형성된 관심과 수요를 얼마나 효율적으로 전환시키는가를 설명하는 요인에 가깝다. 따라서 성능 측면에서도 콘텐츠 및 채널 변수군의 효과는 선별적 이었다.

문장유사율: 0%

에서도 머신러닝이 상위에 차지하고 있고 딥러닝이 뒤따르고 있다. 전반적으로 예측 성능이 개선되는 경향이 있지만 일부 지표(Precision, Recall)와 일부 알고리즘에서는 오히려 성능 감소가 관찰된다.

문장유사율: 22%

평균적으로 Precision과 Recall은 각각 2.84%와 -1.56%의 상승/하락이 있었으며, 이들의 **조화평균인 F1-score 역시 -0.53%로 감소되었다.**

KCI 논문 | 제목 : 청소년기 디지털 ... | 저자 : 장은아(충... | 발행년 : 2025.02

오버샘플링한 자료를 훈련용 및 시험용 자료로 구분하여 XGBoost 분류 모델을 적용한 후 각 패턴별 예측성능을 분석한 결과는 <표 12>와 같다. 분석 결과, 초교 패턴은 모두 정확도(accuracy), 정밀도(precision), 재현율(recall) 모두 0.8~0.9 수준으로 높은 성능을 보이며, 정밀도와 재현율의 **조화평균인 F1 score 역시 0.9에 가깝게 나타났다.**

문장유사율: 0%

유사하게 Accuracy와 AUC도 각각 0.07%와 -1.04%의 상대적 영향을 보였다. 이러한 결과는 채널 효과성 변수가 전반적인 분류 성능을 균일하게 향상시키기 보다는, 특정 지표에 선택적으로 영향을 미쳤음을 시사한다.

문장유사율: 0%

표면적인 지표 하락으로 보일 수 있으나, 이는 예측의 보수성 강화로 재해석 되어야 한다.

문장유사율: 0%

채널 변수는 과거 성공 패턴과 일치하지 않는 모호한 케이스를 Negative로 배제함으로써 정밀도(Precision)를 높이는 필터 역할을 수행 하였다.

문장유사율: 0%

즉, 이 변수군은 구매 파이를 키우기(Recall 증대)보다는, "확실히 살 사람에게만 보낸다"는 효율성 관점의 최적화 기제로 작동하며, 이는 마케팅 비용 절감이 목적인 실무 환경에서 유의미한 트레이드오프 전략으로 활용될 수 있다.

문장유사율: 0%

다시 말해, 해당 변수는 전반적인 반응 탐색 범위를 확장하기보다는, 상대적으로 반응 가능성이 높은 사례를 선별하는데 기여한 것으로 해석할 수 있다.

문장유사율: 0%

종합하면, 마케팅 채널 효과성 변수는 구매 가능성을 폭넓게 포착하는데 기여하기 보다는, 예측 모델이 보다 정제된 기준으로 반응 대상을 선별하도록 유도하는 역할을 수행한다.

문장유사율: 0%

마케팅 채널 효과성 적용에 따른 모델별 성능 4. 지식기반 모델의 예측 성능평가 및 설명력 검증 1) 통합 지식기반 알고리즘별 최종성능 비교는 제안된 4가지 도메인 지식 축을 모두 통합한 고차원 피쳐 공간 하에서, 각 알고리즘의 식별력과 분류 효율성을 종합적으로 검증한 결과이다.

문장유사율: 0%

실험 결과, 머신러닝 기반의 트리 앙상블 모델이 복잡한 심층 신경망 계열 대비 일관된 우위를 보였다.

문장유사율: 0%

우선, 도메인 특화 변수의 도입 효과는 압도적이었다. 전체 알고리즘 평균 기준 Accuracy의 상승폭은 2.54%에 그쳤으나, 핵심지표인 Precision은 9.71%(0.8627 0.9465), Recall은 33.82%(0.5581 0.7468), 그리고 F1-score는 23.14%(0.6766 0.8331) 급증하였다.

문장유사율: 0%

이는 본 연구의 피처 엔지니어링이 단순히 데이터의 차원을 늘린 것이 아니라, 구매의도 판별에 필요한 정보 이득을 구조적으로 극대화했음을 실증한다.

문장유사율: 0%

특히 최소한 구매패턴을 포착하는 민감도가 비약적으로 개선된 점은 실무적 활용가치를 뒷받침한다.

문장유사율: 0%

XGBoost 알고리즘은 모든 지표에서 최고 성능을 달성하였다(F1: 0.9012, AUC: 0.926). 특히 Recall 0.8438과 Precision 0.9670을 동시에 기록한 것은, 과잉 마케팅 비용을 통제하면서도 잠재 고객의 이탈을 방지해야 하는 비용-효익 트레이드 오프를 최적 수준에서 해결했음을 의미한다.

문장유사율: 0%

이는 Gradient Boosting 방식이 잔차 학습을 통해 데이터 내의 미세한 비선형 경계를 정교하게 보정하는 능력에서 기인한 것으로 해석된다.

문장유사율: 0%

Random Forest는 F1-score 0.8866으로 준수한 성능을 보였으나, Recall(0.8203) 측면에서 XGBoost에 소폭 열세를 보였다.

문장유사율: 0%

이는 배깅(Bagging) 기법이 분산 감소를 통해 일반화 성능을 높이는 데는 유리하나, 희소 클래스의 경계면을 예리하게 포착하는 능력은 부스팅(Boosting) 대비 제한적이기 때문에 풀이된다.

문장유사율: 0%

마지막으로, 딥러닝은 AUC 기준으로는 0.97~0.98 수준의 높은 값을 보였으나, 실질적인 분류 지표인 Recall과 F1-score에서는 트리 모델을 넘어서지 못했다.

문장유사율: 0%

특히 CNN과 MLP는 높은 Precision(0.97~0.93)에도 불구하고 낮은 Recall(0.62~0.69)을 기록하여, 확실한 패턴만 구매로 예측하는 보수적 경향을 드러냈다.

출처 표시 문장

문장유사율: 0%

이는 정형 데이터 특유의 불연속적 초평면을 학습하는데 있어, 딥러닝의 연속적 미분 가능함수 보다는 트리 모델의 방식이 더 적합하다는 최근 연구결과(Grinsztajn et al., 2022)와 맥을 같이한다.

문장유사율: 0%

본 연구의 CRM 데이터 분석 환경에서는 지식기반 변수와 XGBoost 조합이 예측 성능과 해석 가능성, 그리고 연산 효율성 측면에서 최적의 해임이 확인되었다.

문장유사율: 0%

이에 본 연구는 실무적으로 XGBoost를 최종 예측 모델로 선정하고, Random Forest를 보조적 검증 도구로 활용하는 하이브리드 앙상블 전략을 사용할 것을 추천한다.

문장유사율: 0%

전체 변수군 반영시 알고리즘별 최종성능 2) 다차원 성능평가 및 지식의 기여도 본 절에서는 4가지 도메인 지식 축이 예측 성능에 미치는 상대적 기여도를 입체적으로 규명하기 위해, 레이더차트 면적(Radar Chart Area)에 기반한 종합평가 지표를 도입하였다.

문장유사율: 14%

이는 Accuracy, Precision, Recall, F1-score, Accuracy, AUC의 5대 핵심지표를 5각형의 축으로 매핑하고 그 내부 면적의 비율을 산출하는 방식으로, 단일 지표의존성을 탈피하여 모델의 정확성(Accuracy), 탐지력(Sensitivity), 균형성(Balance), 판별력(Se-parability)을 포괄하는 종합 성능 지수로 정의된다.

KCI 논문 | 제목: 청소년기 디지털 ... | 저자: 장은아(충... | 발행년: 2025.02

패널 Accuracy Precision Recall F1 Score 초교 0.865 0.838 0.905 0.871 중학교 0.927 0.914 0.942 0.928 <표 12> 각 패널별 전이 패턴 분류 예측성능 다.

KCI 논문 | 제목: 법령 정보 확인 ... | 저자: 박정원(국... | 발행년: 2025.06

구분 Accuracy Precision Recall F1-Score Average 0.669 0.700 0.682 0.665 All-CV 0.053 0.029 0.045 0.061 구분 Average CV ICC 규제 판단 확신도 (0 ~ 100%) 63.82 0.254 0.699 68 규제연구 제34권 제1호 2025년 6월 AUROC는 분류 임계값을 변화시키며 계산된 민감도(True Positive Rate)와 위양성률(False Positive Rate)의 관계를 종합적으로 반영하는 지표로, 1에 가까울수록 예측 성능이 우수함을 의미한다.

문장유사율: 0%

또한 모든 비교군에 대해 동일한 축 순서를 고정적으로 적용하였으므로 상대적 비교 타당성을 갖는다. 이를 통해 각 도메인 지식 축이 예측성능 전반에 기여하는 상대적 영향력을 동일한 기준 하에서 비교하는 것이 가능해진다.

문장유사율: 0%

아울러 교차검증을 위해 5개 지표의 산술평균을 포함하였고, 레이더차트의 우선순위와 동일하게 관찰되었다.

문장유사율: 0%

분석 결과(및), 지식 축들은 예측 성능에 균일하게 기여하지 않았으며, 마케팅 피로도(82.35%) > 시계열 반응(76.15%) > 구매 주기(74.50%) > 채널 효과성(67.87%) 순의 뚜렷한 위계적 구조를 형성하였다.

문장유사율: 0%

도메인 지식 변수 영향력 우선순위 및 최적 모델링 요약 알고리즘별 지식기반 변수 영향력 비교 레이더차트 가장 높은 기여도(Area Ratio=82.35%)를 기록한 피로도 변수군은 구매 여부를 결정짓는 가장 직접적인 근접 결정요인임이 확인되었다.

문장유사율: 0%

이는 최근 일정기간 동안의 접촉 빈도, 누적 노출 강도, 반응 저하 수준과 같은 피로도 관련 지표가 구매 여부를 설명하는데 있어 가장 직접적인 예측 신호로 작용함을 의미한다.

문장유사율: 0%

최적 모델인 RF는 F1-score 0.8534를 달성하며 탁월한 탐지력을 보였다. 이는 최근의 노출 빈도와 반응이력이 고객의 즉각적 구매 의도를 대변하는 강력한 신호이며, 과도한 마케팅 으로 인한 반응 임계점을 포착하는 것이 예측 정확도 확보의 선결조건임을 의미한다.

문장유사율: 0%

다음으로 시계열 변수군은 두 번째로 높은 기여도(Area Ratio=76.15%)를 보였으며, XGBoost 모델에서 F1-score 0.7980의 최적 성능을 기록하였다.

문장유사율: 0%

해당 축은 요일, 시간대, 급여 주기, 분기말과 같은 시간적 맥락 변수를 중심으로 구성되어 있으며, 이는 구매 행동 자체를 직접적으로 유발하기보다는 구매 가능성이 증폭되는 시점을 조절하는 역할을 수행한다.

문장유사율: 0%

즉, 시간적 맥락은 구매 욕구를 직접 창출하기 보다는, 내재된 욕구가 실제 행동으로 발현되는 최적의 타이밍을 결정하는 조절 변수 역할을 수행한다.

문장유사율: 0%

이는 비교적 안정적인 Precision과 중간 수준의 Recall 조합을 통해 구매 확률의 순위를 안정화하는데 기여하는 것으로 해석된다.

문장유사율: 0%

세 번째로, 구매 주기 변수군(Area Ratio=74.50%)은 RF 모델에서 높은 성능을 보였다. 이 결과는 구매 간격, 누적 구매 이력, 재구매 준비도와 같은 장기적 행동 특성이 고객의 기본적인 구매 성향을 형성하는데 기여함을 시사한다.

문장유사율: 0%

이는 고객의 장기적 소비습관과 재구매 사이클이 개별 고객의 기저 구매 확률을 형성하는 구조적 요인으로 작용하기 때문에 상대적으로 낮은 Recall을 보이는 것으로 해석된다.

문장유사율: 0%

임을 시사한다. 그럼에도 불구하고, 전반적인 예측 안정성과 일관성 측면에서는 중요한 기반 변수로 기능한다.

문장유사율: 0%

가장 낮은 면적(Area Ratio=67.87%)을 기록한 채널 변수군은, 전반적인 성능 향상 보다는 조건부 필터로서의 성격이 강하다.

문장유사율: 0%

최적 모델인 XGBoost의 결과(Precision 0.8724 vs Recall 0.6032)에서 보듯, 이 변수군은 구매 가능성이 모호한 대상을 배제하고 확실한 성공 패턴과 일치하는 타겟만을 선별함으로써 마케팅 효율성을 최적화하는 미세조정 역할을 수행한다.

문장유사율: 0%

이는 채널 유형이나 메시지 포맷과 같은 요소가 구매 결정에 직접적인 영향을 미치지 보다는, 이미 형성된 구매 의사에 대해 설득 강도에 보조적 역할을 수행 함을 의미한다.

문장유사율: 0%

즉, 채널 효과성 변수는 단독으로 강력한 예측력을 발휘하기보다는, 다른 핵심 변수들과 결합될 때 그 효용이 극대화되는 구조적 특성을 지닌다.

문장유사율: 0%

본 실험 결과는 도메인 지식기반 예측 모델이 단순한 변수의 결합이 아니라, 피로도에 의한 1차 스크리닝 시점에 의한 확률 증폭 주기에 의한 기저 보정 채널에 의한 효율화라는 단계적 의사결정 프로세스를 반영하고 있음을 실증적으로 규명하였다.

문장유사율: 0%

마케팅에서 보통 구매이력이 가장 강력한 영향력 으로 알려져 있음에도 불구하고, 본 연구에서는 마케팅 피로도나 시간적 맥락이 더욱 높은 영향력으로 작용했다.

문장유사율: 0%

전통적인 마케팅은 누가(Who)에 집중했으나, 현대의 실시간 이커머스 환경에서는 지금(Now)의 상태가 더욱 중요한 결정변수임이 밝혀졌다.

문장유사율: 0%

3) 모델의 설명력 및 변수 중요도 검증: SHAP 분석 본 연구는 전체 지식 변수를 통합한 XGBoost 모델(최적 모델)이 어떠한 변수들의 조합과 상호작용을 통해 고성능을 달성했는지 규명하기 위해, SHAP 기법을 적용하여 사후 해석을 수행하였다.

문장유사율: 0%

SHAP은 게임이론에 기반하여 각 특성이 개별 예측 값에 미치는 한계 기여도(Marginal Contribution)를 산출하는 방법론으로, 변수의 전역적 중요도(Global Importance) 와 영향의 방향성을 동시에 시각화할 수 있다는 장점이 있다.

문장유사율: 0%

는 테스트 데이터 전체에 대한 SHAP Summary Plot을 도식화한 것이다. Y축은 변수 중요도 순(Rank 1~30)으로 정렬되어 있으며, X축의 SHAP Value는 해당 변수가 구매 확률(Log-odds)에 미치는 영향력을 의미한다.

문장유사율: 0%

점의 색상은 변수 값의 크기(Red: High, Blue: Low)를 나타낸다. 구매 예측 설명을 위한 상위 30위 변수 중요도 및 영향 방향성 결과 분석 결과, 상위 5위권 변수는 본 연구가 제안한 "마케팅 피로도 패턴"과 "구매 주기 패턴"이 장악하고 있음이 확인되었다.

문장유사율: 0%

이는 구매가 단순한 노출이 아닌, 고객의 능동적 반응과 심리적 상태에 의해 결정됨을 시사한다.

인용 포함 문장

문장유사율: 0%

"채널별 특정 주제 캠페인 오픈율(30일 내)"전체 1위를 차지한 은 고객이 최근 일정기간 동안 해당 유형의 메시지에 얼마나 반복적으로 노출되고 반응했는지를 반영하는지표이다.

문장유사율: 0%

해당 변수는 SHAP 값이 양(+)의 방향으로 길게 뻗어 있어, 최근 고객이 특정 주제에 대해 반복적으로 반응(Open)했을 때 구매 확률이 비선형적으로 급증함을 보여준다.

문장유사율: 0%

이는 누적된 관심도가 실제 전환의 가장 강력한 선행지표임을 의미한다. 즉, 최근 반응이력이 활발한 고객일수록 구매 확률이 비선형적으로 급증함을 보여준다.

문장유사율: 0%

또한, 주목할 점은 3위 "이메일 경과시간"과 5위 "푸시 경과시간"의 패턴이다. 두 변수 모두 빨간색 점(높은 값=오래 경과시간)이 우측에, 파란색 점(낮은 값=최근 수신)이 좌측에 분포한다.

문장유사율: 0%

이는 메시지를 받은 지 얼마 안된 시점에서는 구매 확률이 억제되다가, 일정 시간이 경과하여 피로도가 해소될수록 구매 가능성이 높아지는 마케팅 휴식기의 필요성을 정량적으로 입증한다.

문장유사율: 0%

그리고 전체 2위인 구매 및 재구매 주기 패턴의 "재구매 준비도"는 고객 고유의 소비 사 이력에 기반한 잠재 욕구를 대변한다.

문장유사율: 0%

해당 플랫폼을 보면, 파란색 점(준비 안 됨)은 0 근처에 머무르나, 빨간색 점(준비됨)은 우측으로 길게 꼬리를 형성한다.

문장유사율: 0%

이는 준비도가 낮은 상태에서는 구매 확률에 미치는 영향이 중립적이지만, 임계치를 넘어서는 순간 구매 전환의 강력한 트리거로 작동하는 비선형적 특성을 보여준다.

인용 포함 문장

문장유사율: 0%

허나 주목할 점은 고정적 환경 맥락을 의미하는 베이스라인 변수인 다양한 등이 상위권 "캠페인 채널, 채널 메시지 유형, 메시지 수신, 이메일 계정"에 다수 포진해 있다는 사실이다.

문장유사율: 0%

그러나 이들은 대부분이산형 변수로서, 고객이 처한 구조적 환경을 설명한다. 예를 들어, 거래 알림(배송/결제 등)은 고객이 이미 서비스에 깊게 관여하고 있음을 나타내는 대리 지표로 작용하여 구매 확률을 높인다. 즉, 베이스라인 변수가 구매가 일어나기 쉬운 환경을 정의한다면, 도메인 지식 변수는 그 환경 속에서 발생하는 동태적 구매 의도를 포착하여 예측의 정밀도를 완성한다.

인용 포함 문장

문장유사율: 0%

"분기말 근접도, 선호 시간대 일치도, 급여일 근접도, 주말 여부" 또한, 중위권에는 와 같은 시계열적 메시지 반응 패턴 변수들이 위치하여, 이미 형성된 구매 의도가 실행으로 옮겨지는 최적의 타이밍을 조율하는 역할을 수행한다.

문장유사율: 0%

한편, 마케팅 채널 효과성 패턴은 상위권 진입이 제한적이었으나, "주제 경과시간, 주제 노출 수"와 같은 변수들이 하위권에서 관찰되어 설득 효율을 보조적으로 제어하는 요인임이 재확인되었다.

문장유사율: 0%

이는 앞선 레이더차트 분석 결과와 일치하는 것으로, 채널 및 콘텐츠 변수가 구매 자체를 결정하기보다는 이미 형성된 구매 의도를 가진 고객에 대한 설득 효율을 보조적으로 제어하는 요인임을 재확인시켜 준다.

문장유사율: 0%

따라서, SHAP 분석은 본 연구의 예측 모델이 최근 상호작용과 피로도 관리(1차 신호)와 기저 구매 성향(2차 신호)을 기반으로, 고정적 환경(베이스라인)과 시점 맥락(시계열)을 통해 확률을 정교화하는 계층적 의사결정 구조를 학습했음을 의미한다.

문장유사율: 0%

이러한 결과는 본 연구에서 제안한 지식기반 변수 설계가 단순한 특성 나열이 아니라, 실제 구매 의사결정 프레임워크의 타당성을 실증적으로 뒷받침한다.

문장유사율: 0%

V. 토의 및 한계 본 연구는 불확실성이 높은 이커머스 고객 행동 예측 문제에 있어, 도메인 지식에 기반한 피쳐 엔지니어링이 모델의 구조적 복잡성을 대체할 수 있는 효율적인 대안임을 실증하였다.

문장유사율: 0%

실험 결과는 (1) 마케팅 피로도(Fatigue)와 시점 요인이 전통적인 구매 이력보다 강력한 예측 인자임을 규명하였고, (2) 정형 데이터 환경에서 트리 기반 모델이 딥러닝 대비 우월한 비용 효율성을 가짐을 입증하였다.

문장유사율: 0%

본 절에서는 이러한 결과가 갖는 이론적, 방법론적, 실무적 함의를 심층적으로 논의한다.

문장유사율: 0%

1. 이론적 기여: 정적 프로파일링에서 동태적 상태 감지로의 전환 기존의 고객관계관리 연구는 주로 RFM이나 인구통계학적 정보에 기반한 정적 프로파일링에 의존해왔다.

문장유사율: 0%

그러나 본 연구의 결과는 특히 피로도와 관여도와 같이 고객의 동태적 상태가 구매 의사결정의 더 즉각적인 결정 변수임을 시사한다.

문장유사율: 0%

가장 높은 기여도를 보인 "마케팅 피로도 패턴"은 자극-반응이론을 확장하여, 고객이 마케팅 자극을 수용하는데 있어 인지적 자원의 한계와 회복 메커니즘이 작동함을 보여주었다.

문장유사율: 0%

특히 SHAP 분석에서 확인된 경과 시간과의 양의 상관관계는, 메시지 수신 후 일정 기간의 휴식기가 주어졌을 때 구매 확률이 비선형적으로 반등함을 입증한다.

문장유사율: 0%

이는 마케팅을 단순히 더 많이 노출하는 것이 아니라, 고객의 인내심이 회복될 때까지 기다리는 것으로 재정의 해야 함을 이론적으로 뒷받침한다.

문장유사율: 0%

따라서 본 연구의 마케팅 피로도 패턴은 사실상 상호작용의 양면성을 포괄한다. 고객의 능동적 관심과 수동적 피로 사이의 균형점을 찾는 것이 핵심일 수 있다.

문장유사율: 0%

2. 방법론적 함의: 정형 데이터에서의 귀납적 편향과 모델 적합성 본 연구는 희소성과 이질성이 높은 CRM 로그 데이터에서 XGBoost와 같은 트리 기반 앙상블 모델이 고도화된 딥러닝보다 우수한 성능을 보임을 확인하였다.

문장유사율: 0%

이는 Grinsztajn et al. (2022)이 주장한 바와 같이, 정형 데이터 특유의 불연속적인 결정 경계를 학습하는데 있어, 신경망의 매끄러운 함수 근사보다는 트리의 축 정렬 분할 방식이 더 적합한 귀납적 편향을 가지기 때문으로 해석된다.

문장유사율: 0%

또한, 도메인 지식 피치가 투입되었을 때 모든 모델의 성능이 비약적으로 향상된 점은, 데이터 중심 AI(Data-centric AI) 관점에서 모델 아키텍처의 고도화보다 데이터 자체의 표현력을 높이는 것이 성능 개선의 한계 효용이 더 큼을 시사한다.

문장유사율: 0%

3. 실무적 함의: 효율적 자원 배분을 위한 의사결정 프레임워크 본 연구결과는 마케터에게 누가(Who)를 넘어 언제(When), 무엇을(What)을 최적화하는 구체적인 가이드라인을 제공한다.

문장유사율: 0%

연구에서 확인된 변수 중요도의 위계에 따라 다음과 같은 3단계 운영전략을 제안한다.

문장유사율: 0%

1단계는 피로도 기반 필터링으로 피로도 변수를 활용하여 최근 과도한 노출로 반응 억제치가 낮아진 고객을 선제적으로 배제함으로써 이탈을 방지한다.

문장유사율: 0%

2단계는 시점 및 주기기반 타겟팅으로 시계열 반응 패턴과 재구매 준비도가 높은 고객을 식별하여, 구매 욕구가 최고조에 달한 타이밍에 메시지를 발송한다.

문장유사율: 0%

3단계는 채널 효율화로 채널 효과성 변수를 활용하여, 구매 가능성이 확인된 고객에게 가장 설득력 높은 채널과 콘텐츠를 매칭함으로써 전환 비용을 최소화한다.

문장유사율: 0%

추가적으로, 본 연구는 제안된 전략의 효용성을 검증하기 위해 비용편익 분석 시뮬레이션을 수행하였다.

문장유사율: 0%

분석 결과, 최적 모델(XGBoost)을 적용할 경우 베이스라인 대비 동일 예산하에서 약 1.15배의 구매 전환 성과(Precision: 0.8401 → 0.9670)를 창출하며, 동일 목표 달성 기준으로는 불필요한 메시지 발송 비용을 약 13.1% 절감할 수 있는 것으로 추산되었다.

문장유사율: 0%

또한, 기존의 무작위 타겟팅(Precision=10%) 방식과 비교할 경우, 약 9.7배 더 높은 타겟팅 적중률을 보여 마케팅 효율성을 획기적으로 개선할 수 있다.

문장유사율: 0%

이는 정밀도의 향상이 마케팅 ROI의 실질적 개선으로 직결됨을 시사하며, 제안된 프레임워크가 단순한 예측을 넘어 비용 효율적인 의사결정 도구로 기능함을 입증한다.

문장유사율: 0%

4. 한계점 본 연구는 마케팅도메인 지식을 4가지 축으로 구조화하고 이를 통해 예측 성능의 비약적 향상과 설명 가능성을 동시에 달성할 수 있음을 입증하였다.

문장유사율: 0%

이는 복잡한 딥러닝 도입에 부담을 느끼는 실무 현장에 경량화된 고성능 모델이라는 합리적 대안을 제시했다는 점에서 의의가 크다.

문장유사율: 0%

그럼에도 불구하고 본 연구는 특징이 커머스 플랫폼의 데이터에 국한되어 있어, 카테고리(예: 식품 vs 가전)나 비즈니스 모델(예: 구독형 vs 단건 구매)에 따른 변수 중요도의 차이를 포괄적으로 검증하지 못했다는 한계를 갖는다.

문장유사율: 0%

향후 연구에서는 다양한 도메인의 데이터셋을 활용한 교차 검증 및 실제 마케팅 캠페인에 모델을 적용하여 ROI 변화를 측정하는 현장 실험이 요구된다.

문장유사율: 0%

또한, SHAP 분석 결과는 변수와 구매 간의 강한 연관성을 시사하지만, 이것이 엄밀한 인과관계를 의미하지는 않는다. 예를 들어, 오픈율이 높아서 구매한 것인지 구매할 마음이 있어서 오픈을 많이한 것인지에 대한 내생성 문제는 향후 인과 추론 방법론을 통해 규명되어야 한다.

인용 포함 문장

문장유사율: 0%

Ⅶ. 결론 본 연구는 라는 근본적 질문에서 출발하여, 이커머스 고객 행동 예측을 위한 도 "데이터의 표현력(Representation)이 모델의 구조적 복잡성을 능가할 수 있는가?"에 인 지식기반의 프레임워크를 제안하고 실증하였다.

문장유사율: 0%

대규모 로그 데이터를 활용한 실험 결과, 도메인이론(피로도, 시점 역동성 등)을 반영한 특성공학은 예측 정확도를 획기적으로 개선 하였으며, 특히 경량화된 XGBoost 알고리즘이 고도화된 시계열 딥러닝 모델을 상회하는 성능을 입증하였다.

문장유사율: 0%

이는 희소성과 이질성이 높은 CRM 데이터 환경에서는 맹목적인 모델 깊이의 추구보다 데이터의 질적 맥락을 보강하는 것이 성공의 핵심 요인임을 시사한다.

문장유사율: 0%

본 연구의 핵심 결론은 다음 세가지로 요약된다. 첫째, 데이터 중심 AI의 유효성을 입증 하였다.

문장유사율: 0%

수억 건의 데이터라도 정보 밀도가 낮다면 딥러닝 모델은 과적합되거나 수렴에 실패할 수 있음을 확인하였다.

문장유사율: 0%

반면, 마케팅도메인 지식을 통해 데이터에 귀납적 편향을 주입했을 때 성능이 비약적으로 향상됨을 증명함으로써, 빅데이터 시대에도 여전히 데이터 과학자의 도메인 이해도가 모델 성능의 상한선을 결정하는 변수임을 확인하였다.

문장유사율: 0%

둘째, 동태적 마케팅 제어 변수의 발견이다. 구매 예측의 결정적 변수는 정적인 인구통계학적 속성이 아니라, 동적인 피로도와 시점 역동성 임이 밝혀졌다.

문장유사율: 0%

이는 마케팅이 고정된 타겟팅에서 벗어나, 고객의 실시간 상태에 따라 메시지 발송 빈도와 타이밍을 조절하는 동적 제어 시스템으로 진화해야 함을 의미한다.

문장유사율: 0%

셋째, 지속가능한 고효율 모델링의 실현이다. 고가의 GPU 인프라를 요구하는 딥러닝 대비, 본 연구의 방법론은 CPU 환경에서도 더 높은 성능과 빠른 추론 속도를 달성했다.

문장유사율: 0%

이는 탄소배출을 저감하는 Green AI의 흐름에 부합하며, 리소스가 제한적인 기업들에게 즉시 도입 가능한 현실적인 고도화 전략을 제시한다.

문장유사율: 0%

결론적으로, 본 연구는 데이터의 희소성이라는 사막 속에서 복잡한 딥러닝이라는 신기루를 쫓는 대신, 도메인 지식이라는 나침반을 장착한 경량 모델이 성능과 설명력, 비용 효율성 모든 면에서 압도적인 우위를 점할 수 있음을 증명한 이론 기반 비즈니스 애널리틱스 프레임워크이다.

문장유사율: 0%

특히 시뮬레이션을 통해 입증된 13.1%의 비용 절감과 1.15배의 성과 향상은 본 연구가 단순한 학술적 탐구를 넘어 기업의 이익을 극대화하는 실질적인 적정 기술로 기능할 수 있음을 강력히 뒷받침한다.

문장유사율: 0%

향후 본 연구자인과 추론과 결합하여 마케팅의 진정한 증분 효과를 측정하는 연구로 확장되기를 기대한다.

참고문헌

참고문헌 Akter, J., Roy, A., Rahman, S., Mohona, S., & Ara, J. (2025). Journal of "Artificial intelligence-driven customer lifetime value (CLV) forecasting: Integrating RFM analysis with machine learning for strategic customer retention," *f Computer Science and Technology Studies*, 7(1), pp.249–257. Althoff, T., Horvitz, E., White, R. W., & Zeitzer, J. (Year). Proceedings of the 26th international "Harnessing the web for population-scale physiological sensing: A case study of sleep and performance," conference on WorldWide Web, Alvarez-Melis, D., & Jaakkola, T. S. (2018). "On the robustness of interpretability methods," arXiv preprint arXiv:1806.08049. Anderl, E., Becker, I., Von Wangenheim, F., & Schumann, J. H. (2016). "Mapping the customer journey: Lessons learned from graph-based online attribution modeling," *International journal of research in marketing*, 33(3), pp.457–474. Arik, S. Ö., & Pfister, T. (Year). "Tabnet: Attentive interpretable tabular learning," Proceedings of the AAAI conference on artificial intelligence,

Attari, V., & Arroyave, R. (2025). *Digital Discovery*, 4(10), pp.2765–2780. Attentive. (2025). Consumer Trends Report: The State of Personalized Marketing in 2025. A. M. Inc. <https://go.attentivemobile.com/hubfs/Consumer%20Trends%20Report%20The%20State%20of%20Personalized%20Marketing%20in%202025.pdf> Bengio, Y., Courville, A. C., & Vincent, P. (2012). CoRR, abs/1206.5538, 1(2) "Unsupervised feature learning and deep learning: A review and new perspectives," 665), p.2012. Berger, P. D., & Nasr, N. I. (1998). "Customer lifetime value: Marketing models and applications," *Journal of interactive marketing*, 12(1), pp.17–30. Beyazit, E., Kozaczuk, J., Li, B., Wallace, V., & Fadlallah, B. (2023). Advancing systems, 36, pp.43108–43135. Bonfrer, A., & Drèze, X. (2009). Marketing "Real-time evaluation of e-mail campaign performance," *Science*, 28(2), pp.251–263.

Chamberlain, B. P., Cardoso, A., Liu, C. B., Pagliari, R., & Deisenroth, M. P. (Year). "Customer lifetime value prediction using embeddings," Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, Chen, K.-Y., Chiang, P.-H., Chou, H.-R., Chen, T.-W., & Chang, T. "Trompt: Towards a better deep neural network for tabular data," -H. (2023). arXiv preprint arXiv:2305.18446. Chen, Q., Zhao, H., Li, W., Huang, P., & Ou, W. (Year). "Behavior sequence transformer for e-commerce recommendation in alibaba," Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data, Daft, R. L., & Lengel, R. H. (1986). *Management Science*, 32(5), pp.554–571. Du, M., Hammer schmidt, C. A., Varisteas, G., State, R., Brorsson, M., & Zhang, Z. (Year). ESANN, EU. (2024). Official Jou "time series modelling of market price in real-time bidding," *Journal of the European Association for Artificial Intelligence*, 2024, pp.1–10. "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence," *Official Journal of the European Union*, p.2024/1689.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). "RFM and CLV: Using iso-value curves for customer base analysis," *Journal of marketing research*, 42(4), pp.415–430. Fader, P. S., Hardie, B. G., Lee, K. L., & Ka Lok, (2005). "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Journal of Marketing Science*, 24(2), pp.275–284. Farjon, G., & Bar-Hillel, A. (2022). IEEE Access, 10, pp.75376–75384. Glikson, E., & Woolley, A. W. (2020). *Academy of Management Annals*, 14(2), pp.627–660. Gorishniy, Y., Kotelnikov, A., & Babenko, A. (2024). arXiv preprint arXiv:2410.24210. Gorishniy, Y., Rubachev, I., & Khoshdel, A. (2024). "Tabm: Advancing tabular deep learning with parameter-efficient ensembling," *Advances in neural information processing systems*, 34, pp.18932–18943. Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). "Why do tree-based models still outperform deep learning on typical tabular data?", *Advances in neural information processing systems*, 35, pp.507–520.

GrowthSuite. Access Provider, <https://www.growthsuite.net/resources/shopify->

"Abandoned Cart Email Timing: When to Send vs When to Stop," *cart-abandonment/timing-strategy*, retrieved January 15, 2026. Hacker, P. (2024). "Sustainable AI regulation," *Common Market Law Review*, 61(2). Harris, P., Pol, H., & Van Der Veen, G. (2020). Customer journey: From practice to theory. In *The Routledge companion to strategic marketing* (pp. 67–90). Routledge. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (Year). recommendation, "ProLightcn: Simplifying and powering graph convolution network forcecedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, He, Z., Xie, Z., Zhu, Q., & Qin, Z. (Year). International C "Sparse doubledescend: Where network pruning aggravates overfitting," conference on Machine Learning, Hollmann, N., Müller, S., Purucker, L., Krishna kumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., & Hutter, F. (2025). "Accurate predictions on small data with a tabular foundation model," *Nature*, 637(8045), pp.319–326.

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). "Tabtransformer: Tabular data modeling using contextual embeddings," *arXiv preprint arXiv:2012.06678*. Hwang, Y. a., & Song, J. (2023). *Communications for Statistical Application* "Recent deep learning methods for tabular data," *tions and Methods*, 30(2), p.215–226. Jannach, D., & Ludewig, M. (Year). *Proceedings of the eleventh ACM "When recurrent neural networks meet the neighborhood for session-based recommendation,* " M conference on recommender systems, Jia, R., Li, R., Yu, M., & Wang, S. (Ye "E-commerce purchase prediction approach by user behavior data," *ar*). 2017 i nternational conference on computer, information and telecommunication systems (CITS), Kang, W.-C., & McAuley, J. (Year). "Self-attentive sequential recommendation," 2018 IEEE international conference on data mining (ICDM), Koren, Y. (Year). "Collaborative filtering with temporal dynamics," *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*,

Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). *California management "Understanding the role of artificial intelligence in personalized engagement marketing,* " ment review, 61(4), pp.135–155. Lambrecht, A., & Tucker, C. (2013). *Journal of "When does retargeting work? Information specificity in online advertising,* " marketing research, 50(5), pp.561–576. Lemon, K. N., & Verhoef, P. C. (2016). "Understanding customer experience throughout the customer journey," *Journal of marketing*, 80(6), pp.69–96. Lindskog, W., & Prehofer, C. (Year). 2023 Eighth "A federated learning benchmark on tabular data: Comparing tree-based models and neural networks," th International Conference on Fog and Mobile Edge Computing (FMEC), Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). "From local explanations to global understanding with explainable AI for trees," *Nature machine intelligence*, 2(1), pp.56–67. Lundberg, S. M., & Lee, S.-I. (2017). *Advances in neural information "A unified approach to interpreting model predictions,* " ion processing systems, 30,

Maileon. "Email marketing trends, 2023," Access Provider, <https://maileon.com/blogs/email-marketing-trends-2023/>, retrieved January 15, 2026. MailerLite. A "Email Marketing Benchmarks 2025: Is Your Open Rate on Track?," *ccess Provider*, <https://www.mailerlite.com/blog/compare-your-email-performance-metrics-industry-benchmarks>, retrieved January 15, 2026. McKinsey & Company "Building AI trust: The key role of explainability," Access Provider, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability>, retrieved January 15, 2026. Olafsson, A., & Pagel, M. (2018). *The Review of Financial Studies*, 31(11), pp.4398–4446. Pfeifer, P. E "The liquid hand-to-mouth: Evidence from personal finance management software," .. Haskins, M. E., & Conroy, R. M. (2005). *Journal of managerial issues*, pp.11– "Customer lifetime value, customer profitability, and the treatment of acquisitions pending," 25. Pielot, M., & Rello, L. (Year). *Proceedings of the 19th International Conference "Productive, anxious, lonely: 24 hours without push notifications," nce on Human-Computer Interaction with Mobile Devices and Services*,

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (Year). "Manipulating and measuring model interpretability," *Proceedings of the 2021 CHI conference on human factors in computing systems*, Ranganathan, C., & Ganapathy, S. (2002). *Information & management*, 39(6) "Key dimensions of business-to-consumer websites,"), pp.457–465. Rejoinder. "When Is the Best Time to Send an Abandoned Cart Email? (And How to Do It Right)," Access Provider, retrieved January 15, 2026. Ribeiro, M. T., Singh, S., & Guestrin, C. (Year). ""Why should I trust you?" Explaining the predictions of any classifier," *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, Salakhutdinov, R. (2015). "Learning deep generative models," *Annual Review of Statistics and Its Application*, 2(1), pp.361–385. Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). *Management science "Counting your customers: Who-are they and what will they do next?," ce*, 33(1), pp.1–24.

Shannon, C. E. (1948). "A mathematical theory of communication," TheBell system technical journal, 27(3), pp.379-423. Shin, J., Kim, W., Lee, J., Park, J., & Ock, C. Y. (Year). 2022 14th International Conference on Knowledge and System "Controlling Weight Update Probability of Sparse Features in Machine Learning,"
s Engineering (KSE), Shwartz-Ziv, R., & Armon, A. (2022). "Tabular data: Deep learning is not all you need," Information Fusion, 81, pp.84-90. Somvanshi, S., Das, S., Javed, S. A., Antarkisa, G., & Hossain, A. (2024). arXiv preprint arXiv: "A survey on deep tabular learning," 2410.12034. Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (Year). Proceedings of the IEEE international conference on "Revisiting unreasonable effectiveness of data in deep learning era," computer vision, Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Inform "Data imbalance in classification: Experimental evaluation," ation Sciences, 513 , pp.429-441. van Tol, D. S. (2024). Evaluating the feasibility of supervised machine learning models for multi-touch attribution Erasmus University Rotterdam].

Vasileva, S. (Year). Competitivitatea și inovarea în economia cunoașterii, Vaswa "Cutting through the noise: understanding the usage of the customer journey mapping,"
ni, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). "Attention is all you need," Advances in neural information processing systems, 30. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). European journal of operational research, 218(1), pp.211 "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach,"
-229. Wang, X., He, X., Cao, Y., Liu, M., & Chua, T.-S. (Year). Proceedings of "Kgat: Knowledge graph attention network for recommendation," f the 25th ACM SIGKDD international conference on knowledge discovery & data mining, Wathieu, L. (2004). "Consumer habituation," Management science, 50(5), pp.587-596. Wydmanski, W., Bulenok, O., & Śmieja, M. (Year). 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Zhang, H., Arvin, C., Efimov, D., Mahoney, M. W., Perrault-Joncas, D., Ramasubramanian, S., Wilson, A. G., & Wolff, M. (2024). arXiv preprint arXiv:2412.02525. Zi "Llmforecaster: Improving seasonal event forecasts with unstructured textual data,"
egler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (Year). Proceedings of "Improving recommendation lists through topic diversification," f the 14th international conference on WorldWide Web,

문장유사율: 0%

관계형 테이블 세부 속성 및 설명 메시지 전송 및 반응 로그 (Interaction Logs) - 속성 : 발송 시각(Timestamp), 채널(Channel), 플랫폼, 메시지 유형, 이메일 제공자 - 반응: Open, Click, Purchase 등 고객의 단계별 행동 데이터 캠페인 메타 정보 (Campaign Metadata) - 속성: 캠페인 유형(Bulk/Trigger/Transactional), 주제(Topic), 마케팅 채널 - 설명: 마케팅 전략 및 운영기간 등 공급자 측면의의도(Intent) 정보 고객 관계속성 (Customer Attributes) - 속성: 고객 식별자(User ID), 최초 구매일(First PurchaseDate) - 설명: 고객의 라이프 사이클(Lifecycle) 단계 및 충성도를 추정하는 기준 정보 외부 시점 정보 (Temporal Context) - 속성: 공휴일 여부, 요일, 월 정보 - 설명: 계절성(Seasonality) 및 사회적 이벤트가 소비에 미치는 외부 효과 정보

문장유사율: 0%

그룹(Group) 변수명(Variable) Mean Std Min Median Max Skewness 종속변수 구매 여부(Target) 베이스라인 채널별 특정 주제 캠페인 오픈율(30일 내)

문장유사율: 0%

캠페인 지속시간(평균) 불만 접수 경과시간(평균) 가입기간(첫 구매 경과) 마지막 클릭 경과시간(평균)

문장유사율: 0%

마지막 오픈 경과시간(평균) 구독 취소 경과시간(평균) 공휴일 여부 총 캠페인 수 총 메시지 수신 수

문장유사율: 0%

총 구매 횟수 메시지 수신: 기타 -2.1524 메시지 수신: 데스크탑 메시지 수신: 패블릿 메시지 수신: 스마트폰

문장유사율: 0%

메시지 수신: 태블릿 발송 채널: 이메일 -0.4631 발송 채널: 모바일 푸시 발송 채널: 웹 푸시 이메일 계정: 지메일이 메일 계정: Mail.ru

문장유사율: 0%

이메일 계정: 기타 -0.2649 채널 캠페인 유형: 전체 유저 발송 -0.7235 채널 캠페인 유형: 기타 거래 발송

문장유사율: 0%

채널 캠페인 유형: 특정 유저행동 발송 캠페인 채널: 이메일 캠페인 채널: 모바일 푸시 캠페인 채널: 멀티채널

문장유사율: 0%

캠페인 채널: SMS 캠페인 주제: 이벤트 캠페인 주제: 생일 축하 캠페인 주제: 리뷰 요청 캠페인 주제: 연관 상품 추천

문장유사율: 0%

캠페인 주제: 기타 캠페인 주제: 할인정보 -0.6835 채널 메시지 유형: 전체 유저 발송 -0.7224 채널 메시지 유형: 기타 거래 발송

문장유사율: 0%

채널 메시지 유형: 특정 유저행동 발송 메시지 클릭 여부 메시지 불만 접수 여부 메시지 오픈 여부

문장유사율: 0%

메시지 구독 취소 여부 구매 및 재구매 마지막 구매 경과일 재구매 준비도 -4.4794 구매 후 불응기

문장유사율: 0%

-4.9744 시계열적 메시지 선호 시간대 일치도 선호 요일 일치도 주말 여부 주차정보 급여일 근접도

문장유사율: 0%

분기말 근접도 마케팅 피로도 마케팅 피로도 이메일 경과시간 푸시 경과시간 마지막 접속 경과시간

문장유사율: 0%

발송 간격 표준편차 유저 오픈율(30일 내) 유저 클릭율(30일 내) 유저 오픈 횟수(30일 내) 채널별 특정 주제 캠페인 오픈율(30일 내)

문장유사율: 0%

마케팅 채널 효과성 주제 신선도 -1.3031 주제 노출 수(7일) 주제 경과시간 성공 경로 일치도

문장유사율: 0%

성공 캠페인 유사도 알고리즘 파라미터 탐색범위 최적값 Random Forest n_estimator s 100 ~ 800 (step=50) 700

문장유사율: 0%

max_depth 10 ~ 30 (step=5) 30 min_samples_split 2 ~ 10 2 min_samples_leaf 1 ~ 5 8 XGBoost n_estimators 100 ~ 1,000 450 learning_rate 0.001 ~ 0.3 (log) 0.0608

문장유사율: 0%

max_depth 3 ~ 10 3 min_child_weight 1 ~ 10 4 subsample 0.5 ~ 1.0 0.5733 colsample_bytree 0.5 ~ 1.0 0.5461 gamma 0 ~ 5 0.9313 reg_alpha 1e-5 ~ 10 (log) 0.0011

문장유사율: 0%

reg_lambda 1e-5 ~ 10 (log) 0.0002 MLP n_layers 1 ~ 3 2 activation relu/gelu/selu gelu units_l0 32 ~ 512 128 dropout_l0 0.0 ~ 0.5 0.2080 l2 1e-6 ~ 1e-2 4.8 7E-03

문장유사율: 0%

optimizer adam/rmsprop/sgd rmsprop learning_rate 1e-4 ~ 1e-2 2.83E-04 batch_size 16/32/64/128 32 CNN conv_filters 32 ~ 256 224 kernel_size 2 ~ 5 2

문장유사율: 0%

pool_size 2 ~ 3 3 num_conv_layers 1 ~ 2 2 dense_units 32 ~ 256 192 dropout_rate 0.1 ~ 0.6 0.2400 l2 1e-6 ~ 1e-2 1.44E-04 learning_rate 1e-5 ~ 1e-3 1.17E-04

문장유사율: 0%

batch_size 32/64/128 32 RNN num_layers 1 ~ 2 2 rnn_units 32 ~ 256 224 rnn_units2 16 ~ 128 128 dense_units 16 ~ 256 96 dropout_rate 0.0 ~ 0.5 0.3461

문장유사율: 0%

l2 1e-6 ~ 1e-2 3.20E-03 activation tanh/relu/sigmoid tanh learning_rate 1e-5
~ 1e-2 1.21E-05 batch_size 32/64/128 32 LSTM num_lstm_layers 1 ~ 2 2

문장유사율: 0%

lstm_units_1 32 ~ 256 96 lstm_units_2 32 ~ 256 112 dense_units 16 ~ 64 48 dr
opout_rate 0.1 ~ 0.5 0.4530 l2 1e-5 ~ 1e-2 7.43E-04 learning_rate 1e-5 ~ 1e-
3 3.18E-04

문장유사율: 0%

batch_size 32/64/128 128 CNN-LSTM conv_filters 16 ~ 128 48 kernel_size 2
~ 5 4 pool_size 2 ~ 4 4 lstm_units 32 ~ 256 256 dense_units 16 ~ 64 16

문장유사율: 0%

dropout_rate 0.1 ~ 0.5 0.1156 l2 1e-5 ~ 1e-2 3.23E-05 learning_rate 1e-5 ~ 1e-
3 5.71E-04 batch_size 32/64/128 128 RNN-LSTM lstm_units1 32 ~ 256 256

문장유사율: 0%

lstm_units2 16 ~ 128 64 dense_units 16 ~ 64 64 dropout_rate 0.1 ~ 0.5 0.4873
l2 1e-5 ~ 1e-2 7.33E-04 learning_rate 1e-5 ~ 1e-3 1.77E-04 batch_size 32/64
/128 32

문장유사율: 0%

Algorithm Precision Recall F1-score Accuracy AUC (ROC) RF XGB MLP CN
N RNN LSTM RNN-LSTM CNN-LSTM Algorithm Ranking Precision Recall
F1-score Accuracy AUC Baseline 구매 및 재구매 주기 패턴 Baseline 구매 및 재구
매 주기 패턴 Baseline 구매 및 재구매 주기 패턴 Baseline 구매 및 재구매 주기 패턴 Ba
seline 구매 및 재구매 주기 패턴 RF XGB RNN-LSTM CNN-LSTM LSTM RNN
MLP CNN Average Gain 0.64% 10.39% 6.34% 0.59% 0.68%

문장유사율: 0%

Algorithm Ranking Precision Recall F1-score Accuracy AUC Baseline 시계열적
메시지 반응 패턴 Baseline 시계열적 메시지 반응 패턴 Baseline 시계열적 메시지 반응
패턴 Baseline 시계열적 메시지 반응 패턴 Baseline 시계열적 메시지 반응 패턴 XGB R
F CNN-LSTM RNN RNN-LSTM LSTM MLP CNN Average Gain 2.08% 9.0
2% 5.15% 0.51% 0.86%

문장유사율: 0%

Algorithm Ranking Precision Recall F1-score Accuracy AUC Baseline 마케팅 피
로도 패턴 Baseline 마케팅 피로도 패턴 Baseline 마케팅 피로도 패턴 Baseline 마케팅
피로도 패턴 Baseline 마케팅 피로도 패턴 RF XGB CNN-LSTM RNN-LSTM LST
M RNN MLP CNN Average Gain 10.28% 31.09% 21.91% 2.41% 1.22%

문장유사율: 0%

Algorithm Ranking Precision Recall F1-score Accuracy AUC Baseline 마케팅 채
널 효과성 Baseline 마케팅 채널 효과성 Baseline 마케팅 채널 효과성 Baseline 마케팅
채널 효과성 Baseline 마케팅 채널 효과성 XGB RF RNN-LSTM LSTM RNN CN
N-LSTM MLP CNN Average Gain 2.84% -1.56% -0.53% 0.07% -1.04%

문장유사율: 0%

Algorithm Precision Recall F1-score Accuracy AUC Baseline Total Baseline Tot
al Baseline Total Baseline Total Baseline Total XGB RF LSTM CNN-LSTM R
NN-LSTM RNN MLP CNN Average Gain 9.71% 33.82% 23.14% 2.54% 1.32
%

문장유사율: 0%

지식 영향력 우선순위 최적알고리즘 Precision Recall F1-score Accuracy AUC 5 Av
erage Radar Chart 마케팅 피로도 패턴 RF 82.35% 시계열적 메시지 반응 패턴 XGB
76.15% 구매 및 재구매 주기 패턴 RF 74.50% 마케팅 채널 효과성 패턴 XGB 67.87%