

Machine Learning-based Learning-to-Rank Approach for Horse Race Prediction and Web Service Development

Yubin So*, Eunbi Woo*, Hanjun Lee**

*Student, Department of Management Information Systems, Myongji University, Seoul, Korea

**Associate Professor, Department of Management Information Systems, Myongji University, Seoul, Korea

[Abstract]

This study presents a Learning-to-Rank (LTR)-based machine learning approach for predicting relative rankings in horse races. Using 9,140 race records from the Korea Racing Authority (May 2024 – April 2025), we evaluate and compare three Gradient Boosted Decision Tree (GBDT) models: LightGBM, XGBoost, and CatBoost. The proposed framework applies the Listwise LambdaRank technique to optimize overall rankings and assesses performance using NDCG, MAP, and MRR metrics. Results show that CatBoost achieved the highest ranking quality (NDCG = 0.8895, MAP = 0.4204), while LightGBM and XGBoost delivered superior prediction accuracy in practical betting scenarios. Feature importance analysis revealed that recent race performance, overall average rank, assigned weight, and horse age are key predictive factors. Finally, we developed a user-friendly web-based prediction system that visualizes ranking results and supports intuitive decision-making, even for novice users. The proposed framework offers a practical and interpretable solution for accurate horse race prediction.

▶ **Key words:** Horse Racing Prediction, Machine Learning, Learning-to-Rank, LambdaRank, Web-based Prediction System

[요 약]

본 연구는 경마 경기 결과 예측을 위해 Learning-to-Rank(LTR) 기반 머신러닝 모델을 적용하여 경주마 간 상대적 순위를 직접 예측하는 방법을 제안한다. 한국마사회에서 제공한 2024년 5월부터 2025년 4월까지의 9,140건 경주 데이터를 활용하여 LightGBM, XGBoost, CatBoost 모델을 비교 분석하였다. Listwise LambdaRank 방식을 적용하여 전체 순위를 최적화하고 NDCG, MAP, MRR 등의 랭킹 지표로 성능을 평가한 결과, CatBoost가 NDCG(0.8895)와 MAP(0.4204)에서 가장 우수했으나, 실제 베팅 시나리오에서는 LightGBM과 XGBoost가 더 높은 적중률을 보였다. 또한 변수 중요도 분석을 통해 직전 5경기 평균 순위, 평균 순위, 부담 중량, 말의 나이 등이 핵심 예측 요인으로 확인되었으며, 이를 기반으로 사용자 친화적인 웹 기반 경마 예측 시스템을 개발하였다.

▶ **주제어:** 경마 예측, 머신러닝, Learning-to-Rank, LambdaRank, 웹 기반 예측 시스템

- First Author: Yubin So, Corresponding Author: Hanjun Lee
- *Yubin So (ybso5062@gmail.com), Department of Management Information Systems, Myongji University
- *Eunbi Woo (woeeunbee@gmail.com), Department of Management Information Systems, Myongji University
- **Hanjun Lee (hjlee1609@gmail.com), Department of Management Information Systems, Myongji University
- Received: 2025. 09. 01, Revised: 2025. 09. 24, Accepted: 2025. 10. 22.

I. Introduction

경마는 전 세계적으로 인기 있는 스포츠로, 경기 결과에 직접 연동되는 금전적 보상 구조로 인해 정확한 순위 예측에 대한 학술적·실용적 관심을 꾸준히 받아왔다[1][2][3][4]. 국내에서도 경마는 하나의 문화로 자리 잡았으며, 특히 20~40세대의 유입이 뚜렷하게 증가하고 있다. 이에 한국마사회는 젊은 층을 위한 신규 관객 유입 전략을 적극적으로 추진하고 있다. 경마 경기 결과는 경주마의 과거 성적, 기수, 부담중량, 날씨, 트랙 상태 등 다양한 변수의 복합적이고 비선형적인 상호작용에 의해 결정된다. 이러한 복잡성은 초보자가 기존의 전통적 예상 자료를 해석하고 활용하기 어렵게 하며, 데이터 기반 전략적 판단을 저해하는 요인으로 작용한다. 따라서 초보 사용자도 직관적으로 이해하고 활용할 수 있는 예측 시스템의 필요가 크다.

이러한 배경에서, 지금까지 다양한 경마 순위 예측 모델 연구가 수행되어 왔다. 기존 연구들은 한국마사회 데이터를 활용하여 로지스틱 회귀, 랜덤 포레스트, 선형 회귀 기반 모델을 적용하고, 단승식·복승식·삼복승식 예측률을 주요 평가 지표로 활용하였다. 마체중, 우승 비율, 출발 위치 등 주요 변수의 영향력이 분석되었으나, 대부분은 개별 경주마의 승리 확률 또는 기록 예측에 중점을 두었으며, 경주마 간 상대적 순위를 직접 모델링하는 데에는 한계를 보였다[1].

이 한계를 극복하기 위해, 최근에는 Learning-to-Rank(LTR) 기반 연구가 활발히 진행되고 있다[2]. 기존 LTR 연구들은 Pointwise, Pairwise 방식을 적용해, RankNet·LambdaMART 기반의 XGBoost·LightGBM·CatBoost Ranker 등을 비교했으며, Pairwise 방식이 Pointwise 방식보다 우수한 성능을 보였고, 특히 CatBoost Ranker가 모든 평가 지표에서 뛰어났다[2]. 그러나 Pairwise 방식은 두 개체 간 비교에 국한되어 전체 경주마 집합의 순위를 일괄 학습하기 어렵다는 구조적 한계가 있다.

본 연구는 이러한 구조적 한계를 보완하고자 전체 순위를 직접 고려하는 Listwise 방식의 LambdaRank 알고리즘을 LightGBM에 적용하였다. 또한 XGBoost와 CatBoost의 성능 비교로 LTR 방식 간 차이를 분석하고, 변수 중요도 분석을 통해 주요 예측 요인을 파악하였다. 마지막으로 초보자도 쉽게 활용할 수 있도록 예측 결과를 웹 기반 대시보드로 시각화하여 접근성을 높였다.

본 논문의 구성은 다음과 같다. 2장에서는 한국마사회 경마 데이터를 기반으로 주요 변수 정의, 파생 변수 생성,

전처리 과정을 포함한 탐색적 데이터 분석을 다룬다. 3장에서는 Pairwise 및 Listwise 구조에 맞춘 학습 전략을 제시하고, LightGBM·XGBoost·CatBoost를 활용한 LTR 모델 학습 및 실험 설계를 설명한다. 4장에서는 NDCG, MAP, MRR 등의 랭킹 기반 평가 지표를 활용하여 모델별 성능을 비교 분석하고, 승식별 예측 정확도를 검증한다. 5장에서는 예측 결과를 사용자에게 직관적으로 제공하기 위한 웹 기반 경마 예측 시스템의 설계 및 구현 과정을 다룬다. 끝으로 6장에서는 연구 결과를 요약하고, 모델별 활용 가능성을 제시하며, 향후 연구 방향을 논의한다.

II. Preliminaries

1. Problem Definition and Betting Terminology

본 연구에서는 경마 순위 예측을 순위화 문제(ranking problem)로 재정의하였다. 이에 따라 한 경주에 출전하는 말의 집합을 하나의 그룹으로 정의하고, 상대 순위를 학습하는 방식으로 접근하였다.

경마의 전통적 베팅 방식은 예측의 목표와 밀접한 관련이 있으며, 주요 유형은 단승식, 연승식, 복연승식, 복승식, 쌍승식, 삼복승식 등이 있다. 단승식은 1등으로 도착할 말 1두를 적중시키는 방식이다. 연승식은 1-3등 안에 들어올 말 1두를 적중시키는 방식이며, 출두 마필이 7두 이하인 경우 2등 이내를 대상으로 한다. 복연승식은 1-3등 안에 들어올 말 2두를 순서에 상관없이 적중시키는 방식이다. 복승식은 1등과 2등으로 들어올 말 2두를 순서에 상관없이 적중시키는 방식이다. 쌍승식은 1등과 2등으로 들어올 말 2두를 순서대로 적중시키는 방식이다. 삼복승식은 1-3등으로 들어올 말 3두를 순서와 무관하게 적중시키는 방식이다.

2. Learning to Rank (LTR)

LTR은 검색 엔진, 추천 시스템, 문서 정렬, 스포츠 경기 예측 등 다양한 분야에서 활용되는 머신러닝 방법론으로, 입력 데이터 집합을 특정 기준에 따라 정렬하는 모델을 학습한다[5][6]. Table 1과 같이 LTR은 학습 단위에 따라 크게 세 가지 방식으로 구분된다.

Table 1. Comparison of LTR Methods

Type	Description	Feature
Pointwise	Predict single item	Predict scores for individual items and sort them
Pairwise	Predict relative order of two items	Learn the relative predominance of two items of two items
Listwise	Predict order of full list	Optimize the ranking of the entire list

2.1. Learning Algorithm

Pointwise 방식은 각 항목에 대해 독립적으로 예측값을 산출하고, 이를 기반으로 정렬한다. 구현이 간단하지만 항목 간 상호작용을 반영하기 어렵다. Pairwise 방식은 한번에 2개의 문서를 고려해 비교하는 방식이다. 문서를 하나의 'pair'로 이용한다. Listwise 방식은 전체 리스트를 대상으로 순위를 직접 최적화하는 방식으로, 최종 평가 지표와 더 밀접하게 연결된다. 두 항목 간 비교를 통해 상대적 우위를 학습한다.

2.2. Ranking Model

Ranking Model에는 크게 RankNet, LambdaRank, LambdaMART가 존재한다[7]. RankNet은 Pairwise 접근 방식을 사용하여 두 문서의 상대적인 순위 관계를 학습하는 방식이다. LambdaRank는 정보 검색의 랭킹 학습(Learning to Rank) 기법으로, 기존 Pairwise의 손실 함수 대신 랭킹 평가 지표의 변화량을 직접 그래디언트로 활용한다. LambdaMART는 LambdaRank의 아이디어와 MART(Multiple Additive Regression Trees)의 Gradient Boosting이 결합된 알고리즘이다. LambdaMART의 랭킹 평가 지표로 NDCG와 MAP 등이 있다. NDCG(Normalized Discounted Cumulative Gain)는 랭킹 추천 분야에 활용되는 지표로 모델이 얼마나 관련성에 맞게 추천하였는지 보여주는 수치이다[8]. 상위 순위의 정확도가 큰 영향을 미치며 1에 가까울수록 우수한 모델이라 할 수 있다. LambdaMART는 최적화를 위해 정렬 오류의 기울기(Lambda)를 활용하는 알고리즘이다. 본 연구에서는 LightGBM의 rank_xendcg 목적 함수를 통해 LambdaMART를 구현하였으며, 비교 대상으로 XGBoost와 CatBoost를 함께 실험하였다. 이를 정리하면 Table 2와 같다.

Table 2. Comparison of LTR Models and Objective Functions

Model	Objective Function	LTR Method
LightGBM	rank_xendcg	Listwise
XGBoost	rank_pairwise	Pairwise
CatBoost	YetiRankPairwise	Pairwise

본 연구에서는 이와 같이 서로 다른 LTR 방식을 비교함으로써, 순위 예측 목적에 적합한 모델 구조를 탐색하였다.

3. Evaluation Metrics for Ranking

경마 예측에서는 순위의 전반적인 품질을 평가하는 것이 중요하다. 단순 정확도만으로는 순위 전체의 적합도를 충분히 반영할 수 없으므로, 본 연구에서는 Table 3의 랭킹 기반 지표를 활용하였다.

Table 3. Evaluation Metrics

Metric	Description
NDCG	Measures the quality of a ranked list, assigning higher weights to relevant items at higher positions. Ranges from 0 to 1, with 1 indicating a perfect ranking.
MAP	Evaluates how well relevant items are placed at the top of the ranked list, incorporating both precision and recall.
MRR	Calculates the average of the reciprocal ranks of the first relevant item in the results.
ExactRank Accuracy	Measures the proportion of items whose predicted rank exactly matches the actual rank (specifically defined for this study).

이러한 지표들은 각기 다른 베팅 전략(단승, 복승, 삼복승 등)과 대응될 수 있으며, 다양한 지표를 병행 평가함으로써 모델의 종합적인 성능을 분석하였다.

III. The Proposed Scheme

1. Data Collection and Preprocessing

본 연구는 한국마사회에서 제공하는 경주마·기수·조교사 데이터를 활용하였다. 국내 경마는 서울, 부산, 제주에서 개최되며, 그중 매출액의 55%를 차지하는 서울 경마장 데이터를 대상으로 하였다. 2024년 5월부터 2025년 4월까지 1년간을 대상으로 경주마, 기수, 조교사 정보를 담고 있는 3개의 파일을 각각 수집하고 이를 결합하여 총 9,140건의 관측치를 구성하였다.

경주마, 기수, 조교사 데이터는 기수번호와 조교사 번호를 기준으로 좌측 조인으로 병합하여 메인 데이터 프레임을 구축하였다. 중복되거나, 불필요한 기타 컬럼들은 삭제하였다. 병합 후 발생한 NaN 값에 대해서는 특성에 맞게 0 또는 평균값 등으로 채웠다. 또한 기존에 존재하던 변수 이외에 변수들의 특성을 활용한 추가 파생 변수를 활용하였다.

본 연구에서는 말의 경주 성능을 예측하기 위하여 다양한 파생 변수들을 활용하였다. 먼저, 말의 최근 5경기 평균 순위를 나타내는 '직전_5경기_평균_순위(prev_5_avg_rank_in_race)' 변수와 함께, 말과 기수의 호흡을 반영하는 말-기수 조합 승률 '말_기수_승률(jk_hr_win_rate)'과 '말_기수_평균_순위(jk_hr_avg_rank_in_race_exp)'를 고려하였다. 또한, 조교사의 능력을 나타내는 '조교사 승률(tr_win_rate)'과 '조교사 평균 순위(tr_avg_rank_in_race_exp)' 역시 중요한 변수로 사용되었다.

경주 환경적 요인으로는 '주로의 습도 비율(moisture_percent)'과 '주로 상태를 인코딩한 값(track_condition_category_encoded)', 그리고 '출발 게이트 번호(gate_number)'가 포함되었다. 말의 생체 리듬과 관련된 변수로는 이전 경주로부터의 '휴식 일수(rest_days)'와 경주 시점 '말 나이(age)'가 있었다.

말의 전반적인 경력과 경험을 나타내기 위하여 '말의 총 출전 횟수(game_attendance)', '말의 평균 순위(avg_ord)', 그리고 '말의 승률(1위) (avg_winRate)'을 활용하였다. 계절적 영향을 반영하기 위해서는 '봄 시즌 여부(season_spring)', '여름 시즌 여부(season_summer)', '가을 시즌 여부(season_autumn)', '겨울 시즌 여부(season_winter)'를 원핫인코딩하여 사용하였다.

말의 누적 입상 기록과 관련하여 최근 1년간의 성과를 반영하기 위하여 '1년간 1위 비율(total_ord1CntY)'과 '1년간 2위 비율(total_ord2CntY)'도 계산하여 활용하였다.

말을 고유하게 식별하기 위한 '말 개체 번호(hrNo_category)'와 '출주 번호 범주(chulNo_category)'는 범주형 변수로 처리하여 모델 학습에 사용되었다. 마지막으로, 기수와 조교사의 세부적인 과거 성적을 반영하기 위하여 '기수의 누적 1위 횟수(ord1CntT_jk)', '기수의 누적 2위 횟수(ord2CntT_jk)', '기수의 누적 3위 횟수(ord3CntT_jk)'와 '조교사의 누적 1위 횟수(ord1CntT_tr)', '조교사 누적 2위 횟수(ord2CntT_tr)', '조교사 누적 3위 횟수(ord3CntT_tr)'가 변수로 포함되었다.

이상의 변수들에 대한 기술통계량은 Table 4에서 보는 바와 같다.

Table 4. Descriptive Statistics of Variables

Variable	Mean	Min	Max	Standard Deviation
prev_5_avg_rank_in_race	4.31	0.0	12.0	2.85
jk_hr_win_rate	0.02	0.0	1.0	0.13
jk_hr_avg_rank_in_race_exp	5.70	1.0	12.0	1.67
tr_win_rate	0.10	0.0	1.0	0.06
tr_avg_rank_in_race_exp	5.44	1.0	11.0	0.80
moisture_percent	0.08	0.03	0.2	0.05
track_condition_category_encoded	1.68	0.0	4.0	1.68
rest_days	36.07	0.0	351.0	33.04
age	3.57	2.0	11.0	1.24
game_attendance	2.88	0.0	20.0	2.58
avg_ord	4.33	0.0	16.0	2.93
avg_winRate	0.08	0.0	1.0	0.20
season_spring	0.29	0.0	1.0	0.45
season_summer	0.26	0.0	1.0	0.44
season_autumn	0.27	0.0	1.0	0.44
season_winter	0.18	0.0	1.0	0.39
ord1CntT_jk	379.4	17.0	2246.0	501.2
ord2CntT_jk	361.4	18.0	2097.0	419.35
ord3CntT_jk	344.8	23.0	1793.0	349.41
ord1CntT_tr	326.6	1.00	624.0	177.20
ord2CntT_tr	318.8	1.0	638.0	169.57
ord3CntT_tr	316.8	4.0	622.0	167.89

2. LTR Model Training

LTR은 검색 결과나 추천 시스템에서 항목들의 최적 순위를 학습하는 데 사용되는 머신러닝 방법론이다. 모델은 특정 경주의 모든 출전마의 변수를 입력받아 각 말에 대한 순위 점수를 출력하며, 이 점수를 기준으로 최종 순위를 예측한다. 대표적인 GBDT(Gradient Boosted Decision Tree) 기반 LTR 모델인 LightGBM, XGBoost, CatBoost를 선정하여 비교 분석하였다. 머신러닝 기반 예측 기법은 다양한 분야에서 활용되고 있으며, 본 연구에서도 이러한 방법론을 적용하였다[9].

모델의 일반화 성능을 객관적으로 평가하기 위해 K-Fold 교차검증($n_splits=5$)을 적용하였다. 이때 rcDate와 rcNo를 결합해 각 경주를 고유하게 식별하도록 group_id 변수를 생성하였고, 동일한 경주(group_id) 내의 말들이 분할되지 않도록 group_id 단위로 폴드를 나눔으로써 데이터 누수를 엄격히 방지하였다. 각 폴드는 전체 191개 경주를 약 20%씩 균등하게 포함하도록 구성되었으며, 폴드 간 데이터 중복이 발생하지 않도록 설계하였다. 각 폴드의 훈련 데이터셋으로 모델을 학습시키고, 검증 데

이터셋으로 성능을 평가하였으며 과적합 방지를 위해 조기 종료(early stopping) 기법을 사용하였다.

3. Evaluation

총 191개 경주에 대한 모델별 평균 성능을 계산하였다. 성능평가에서는 NDCG, MAP, MRR을 이용하여 모델 간 성능을 비교하였다. NDCG는 검색 결과의 순서와 각 항목의 연관성 수준을 모두 고려하는 지표로, 얼마나 모델이 적절히 순위를 배치하였는가를 나타내며 0과 1 사이의 값에서 1에 가까울수록 좋다. MAP는 검색 결과의 각 순위별 정밀도를 계산하고, 해당 순위까지의 평균 정밀도(AP)를 구한다. 각 쿼리마다 AP를 계산하고, AP들의 평균이 최종 MAP 값이다. MAP 또한 값이 1에 가까울수록 좋다. MRR은 각 쿼리(질의)에 대한 검색 결과 중 가장 관련성 높은 항목이 나타난 순위의 역수(Reciprocal Rank)를 구한 값들의 평균으로, 값이 1에 가까울수록 좋다. Table 5에서 보는 바와 같이 Catboost의 NDCG 값이 0.8895로 가장 관련성에 맞게 예측한다고 해석할 수 있다. 이는 동일한 패키지 모델로 진행되었던 0.7149[2]와 비교했을 때, 약 0.170 수준으로 성능이 개선되었다. MAP 부분에서도 0.4204로 우수한 성능을 보였다. MRR 값에서는 XGboost가 0.3943으로 우수한 성능을 보였다. 한편, CatBoost의 MRR 값(0.1785)은 상대적으로 낮게 나타났는데, 이는 MRR이 가장 높은 순위의 정답 항목(Top-1)에 민감한 지표이기 때문이다. 본 연구의 Listwise 기반 모델은 Top-1보다는 전체 순위의 일관성 향상을 목표로 학습하기 때문에, Top-k 수준의 순위 품질(NDCG, MAP)은 높았으나 절대 1위 적중률은 낮게 나타나 MRR이 상대적으로 낮게 산출된 것으로 해석된다. 실제 경기에 대한 예측률은 LightGBM이 0.1425로 우수한 예측력을 보였다. 그러나 모델별 성능에서 압도적인 성능 차이는 없었고 평가 지표에 따라 모델의 성능 순위가 달라졌다.

각 성능지표(NDCG, MAP, MRR, ExactRank Accuracy)는 5-Fold 교차검증 결과의 평균값으로 산출되었다. Fold 간 편차는 비교적 작았으며, 모든 모델에서 표준편차가 ± 0.02 이내로 나타나 전반적인 성능 일관성이 확인되었다. 모델 간 평균 성능의 차이는 95% 신뢰수준에서 통계적으로 유의하지 않았으며, 이는 세 모델이 유사한 예측 안정성을 보였음을 의미한다.

Table 5. Performance Test

Model	LightGBM	XGBoost	CatBoost
Average NDCG	0.8725	0.8789	0.8895
Average MAP	0.2263	0.2291	0.4204
Average MRR	0.3940	0.3943	0.1785
ExactRank Accuracy	0.1425	0.1374	0.0768

다음으로, 승식에 따른 모델별 예측 성능을 비교하였다. 단승식은 1위로 들어올 경주마 한 마리를 정확하게 맞히는 방식이다. 연승식은 1위부터 3위 안에 들어올 말 한 마리를 맞히는 방식으로 단승식보다 적중 확률이 높다. 복승식은 1위와 2위로 들어올 경주마 두 마리를 순서에 상관없이 맞히는 방식이며 두 마리를 모두 맞춰야 한다. 쌍승식은 1위와 2위로 들어올 경주마 두 마리를 정확한 순서대로 맞히는 방식으로 복승식보다 난이도가 높지만 큰 배당을 기대할 수 있다. 복연승식은 1위, 2위, 3위 안에 들어올 경주마 두 마리를 순서와 관계없이 맞히는 방식이고 복승식과 연승식을 조합한 형태이다. 삼복승식은 1위, 2위, 3위로 들어올 경주마 세 마리를 순서에 상관없이 맞히는 방식이다. 마지막으로, 삼쌍승식은 1위, 2위, 3위로 들어올 경주마 세 마리를 정확한 순서대로 맞히는 방식으로 가장 난이도가 높다. 비교결과는 Table 6에서 보는 바와 같다. 전반적으로 LightGBM과 XGboost가 우수한 예측능력을 보였다. 단승, 복승의 경우 LightGBM과 XGboost가 각각 18.85%, 7.85%로 우수한 성능을 보였다. 쌍승과 삼쌍승에서 LightGBM이 5.76%와 1.05%로 가장 높은 예측도를 보였다. 특히 가장 어려운 삼쌍승에서 유일하게 1% 이상의 적중률을 기록하였다. 연승, 삼복승, 복연승에서 XGBoost가 각각 52.88%, 5.24%, 19.37% 가장 높은 예측도를 기록하였다. XGBoost는 예측 상위 N위 내에 실제 적중마가 '포함되는' 방식의 승식(예: 연승, 삼복승)에서 가장 뛰어난 성능을 보였다.

모델의 예측 성능을 실제 베팅 상황에 가깝게 평가하기 위해, 예측 점수를 기준으로 상위 1~3위를 선택하는 Top-N 방식의 베팅 시뮬레이션을 수행하였다. 각 선택마에 동일 금액(1단위)을 베팅하고, 실제 배당금(테이크아웃 15%)을 반영하여 ROI(수익률)과 적중률을 계산하였다.

Table 6. Performance Test by Betting Pool

Betting pool	LightGBM	XGBoost	CatBoost
Win	18.85%	18.85%	3.14%
Place/Show	47.12%	52.88%	14.66%
Quinella	7.85%	7.85%	0.00%
Exacta	5.76%	3.14%	0.00%
Trio	3.14%	5.24%	0.00%
Quinella Place	17.28%	19.37%	1.05%
Trifecta	1.05%	0.52%	0.00%

각 모델의 학습과정에서 모델 최적화에 사용된 하이퍼 파라미터들의 값은 Table 7에서 보는 바와 같다.

Table 7. Hyperparameters

Hyperparameter	LightGBM	XGBoost	CatBoost
Objective	rank_xendcg	rank:pairwise	YetiRankPairwise
Metric	ndcg	ndcg	ndcg
Learning_rate	0.05	0.05	0.05
Tree Complexity	num_leaves=31	max_depth=6	depth=6
Regularization	lambda_l2=default		l2_leaf_reg=3
Verbose	-1		0
Data Sampling	subsample=default	subsample=0.8, colsample_bytree=0.8	subsample=default
Early Stopping	callbacks=[lgb.early_stopping(stopping_rounds=10)]		early_stopping_rounds=10
Iteration	num_boost_round=200		iteration=200

4. Feature importance

어떠한 변수들이 실질적인 성능에 영향을 미쳤는지 파악하기 위해 경기의 1위를 예측하는 '단승식'의 변수 중요도 분석을 진행하였다. 변수 중요도는 복잡한 머신러닝 모델의 해석 가능성을 높여, 모델이 어떻게 의사결정하는지 파악하는 데 필수적인 지표이다[10][11]. 분석결과는 Fig. 1에서 보는 바와 같다. 직전 5개 경기에서의 평균 순위 (prev_5_avg_rank_in_race), 평균순위(avg_ord), 부담 중량(wgBudam), 말의 나이(age)의 순으로 중요 변수가 식별되었다.

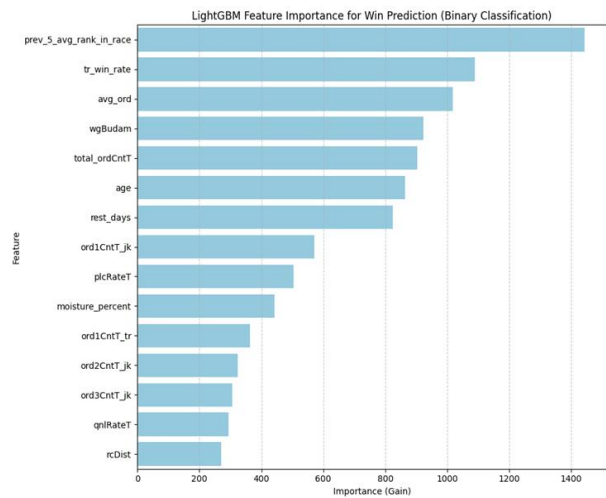


Fig. 1. Feature Importance Analysis

본 연구의 결과는 모델 선정에 있어 '예측 목표'의 중요성을 강조한다. 만일 '학술적 랭킹 품질'이 최우선 목표라면 CatBoost 모델이 가장 적합하다고 볼 수 있다. CatBoost는 NDCG 및 MAP 지표에서 다른 모델들보다 우위를 보이므로, 랭킹 관련성 자체를 가장 잘 파악하는 모델로 평가할 수 있다. 하지만 '실질적인 경마 베팅 적중률'이 최우선 목표라면 LightGBM 또는 XGBoost 모델이 상황에 따라 더 적절하다. 정확한 순서 조합이 중요한 쌍승, 삼쌍승의 적중률을 중요시한다면 LightGBM이 가장 우수한 성능을 보였다. 반면에 Top-N 내 포함 여부 연승, 삼복승, 복연승 적중률을 중요시한다면 XGBoost가 가장 우수한 성능을 보였다.

IV. A System Development of Horseracing Prediction

본 연구에서는 승식에 따라 말들의 조합을 예측하는 웹 서비스를 제안한다. 웹 표준 언어인 HTML5, CSS3, 그리고 JavaScript를 기본으로 사용한다. 사용자 인터페이스의 복잡성과 상호작용성을 고려하여 React, Vue.js, 또는 Angular와 같은 단일 페이지 애플리케이션(SPA) 프레임워크 중 React를 선택하고, 백엔드 개발에는 복잡하고 대규모의 웹 애플리케이션 개발에 적합하며, ORM(Object-Relational Mapping), 관리자 패널 등 다양한 기능이 내장되어 있어 신속한 개발이 가능한 Django를 사용한다. 마사회에서 제공된 데이터를 담은 데이터 베이스로는 Django와 함께 널리 쓰이는 PostgreSQL를 활용하고, 저장소는 AWS S3를 사용한다. 웹은 사용자가 특정 경

기의 날짜, 경주 번호, 출전 말 등의 정보를 검색할 수 있도록 페이지를 제공한다. 백엔드로부터 전달받은 모델의 예측 순위, 각 승식별 적중률 등을 그래프나 표 형태로 직관적으로 보여준다. Fig. 2는 시스템 아키텍처를 나타낸다.

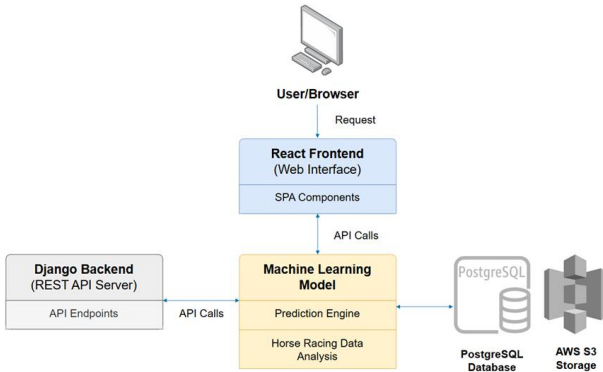


Fig. 2. System Architecture

본 시스템을 통하여 사용자는 날짜, 경마장, 말 이름, 기수 등의 정보를 접근하고, 경주 성적과 배당률 등을 확인할 수 있다. 또한, 본 시스템은 과거 경주 기록 검색 및 통계 시각화 기능을 제공함으로써 사용자의 의사결정을 지원한다. Fig. 3은 예시 화면을 보여준다.

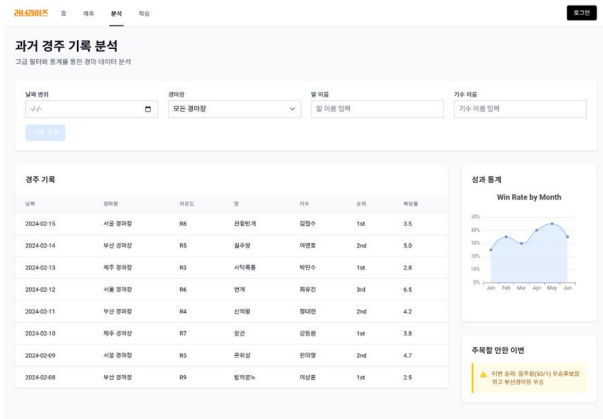


Fig. 3. Interface for Previous Race Analysis

Fig. 4는 본 연구에서 개발한 모델을 활용한 ‘AI 경마 예측’ 서비스 페이지의 모습이다. 사용자는 이곳에서 경기 날짜를 설정하고 출전마들의 이전 경기들의 성적과 머신러닝 모델이 분석한 승리 예측 정보를 제공한다. 또한 승식에 맞게 어떠한 승식에 배팅하는 것이 승리 확률이 높은지 추천 조합과 확률을 제안한다. 사용자는 해당 모델이 추천하는 승식과 경주마 정보를 확인하고, 최적의 조합과 금액으로 서비스를 이용할 수 있다. 또한 승률 임계값을 사용자 임의로 지정할 수 있어 안전추구형, 적극참여형 등 배팅 성향에 맞춘 개인화된 서비스를 경험할 수 있다.

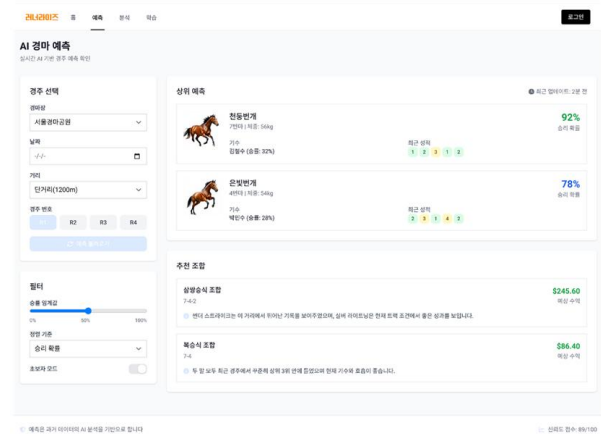


Fig. 4. Web-based Interface for Horse Race Prediction

V. Conclusion

본 연구는 경마 경기 결과 예측을 위한 효과적인 머신러닝 접근법으로 LTR 기반 알고리즘을 적용하고, 다양한 모델 간의 성능을 정량적으로 비교하였다. 특히 경주마 간의 상대적 순위를 직접 고려하는 Listwise 방식의 LambdaRank 알고리즘을 LightGBM 모델에 적용하고, Pairwise 방식의 XGBoost, CatBoost와의 성능 차이를 실험을 통해 얻을 수 있었다.

실험 결과, 모델의 성능은 예측 목표의 성격에 따라 달라지는 양상을 보였다. LightGBM은 단승(18.85%), 쌍승(5.76%), 삼승(1.05%) 등 정확한 순서 조합이 요구되는 승식에서 높은 적중률을 보이며 실용적인 모델로 평가되었다. XGBoost는 연승(52.88%), 복연승(19.37%) 등 Top-N 내 포함 여부를 중시하는 승식에서 가장 뛰어난 예측 성능을 보여 다소 완화된 예측 목표에 부합하는 모델로 나타났다. CatBoost는 NDCG(0.8895), MAP(0.4204) 등 랭킹의 전반적인 품질과 관련성 예측 지표에서 가장 높은 성능을 기록하였으나, 실제 배팅 승식 예측에서는 낮은 적중률을 보여 실용적 순위 예측에는 한계가 있었다.

종합적으로, 예측 목표가 학술적 랭킹 품질 향상이라면 CatBoost가 적합하지만, 본 연구의 핵심 목적이 실제 순위 예측과 배팅 전략의 실용화에 있는 만큼, LightGBM(rank_xendcg 기반)이 가장 효과적인 모델로 판단된다. 특히, 예측이 가장 어려운 삼승에서도 유의미한 적중률을 보였다는 점에서, 초보 사용자도 활용 가능한 사용자 친화적 예측 시스템 개발에 효과적으로 기여할 수 있을 것으로 기대된다.

한편, 본 연구는 단일 연도 데이터와 공개된 정형 변수에 기반해 모델을 구성하였기에, 실제 경주 당일 컨디션이

나 날씨 등 비정형 요인을 반영하지 못했다는 한계가 존재한다. 향후에는 다년간 시계열 데이터 활용, 실시간 정보 기반 예측, 개인화된 전략 추천 시스템, 지표 간 통합 평가 방법론 등의 방향으로 연구를 확장할 수 있을 것으로 기대된다.

REFERENCES

- [1] H. Choe, N. Hwang, C. Hwang, and J. Song, "Analysis of Horse Races: Prediction of Winning Horses in Horse Races Using Statistical Models," *The Korean Journal of Applied Statistics*, Vol. 28, No. 6, pp. 1133-1146, Oct. 2015. DOI: 10.5351/KJAS.2015.28.6.1133
- [2] J. Chung, D. Shin, S. Hwang, and G. Park, "Horse Race Rank Prediction Using Learning-to-rank Approaches", *The Korean Journal of Applied Statistics*, Vol. 37, No. 2, pp.239-253, April 2024. DOI: 10.5351/KJAS.2024.37.2.239
- [3] D. H. Noh, "A Prediction Model for Winning Rate of the Motorboat Racing using Machine Learning," Master's Thesis, Soongsil University, Dec. 2023. <https://www.riss.kr/link?id=T16907201>
- [4] S. J. Kim, "The forecasting model development using data mining analysis of the leisure sports industry : centering horse racing' case," Master's Thesis, Yonsei University, Dec. 2013. <https://www.riss.kr/link?id=T13443618>
- [5] T. Y. Liu, "Learning to Rank for Information Retrieval," *Foundations and Trends in Information Retrieval*, Vol. 3, No. 3, pp. 225-331, Jun. 2009. DOI: 10.1561/1500000016
- [6] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," Cambridge University Press, 2008. <https://nlp.stanford.edu/IR-book/>
- [7] C. J. C. Burges, T. Shaked, T. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to rank using gradient descent," *Proceedings of the 22nd International Conference on Machine Learning*, pp. 89-96, Aug. 2005. DOI: 10.1145/1102351.1102363
- [8] O. Baker, and Q. Yuan, "Machine Learning: Factorization Machines and Normalized Discounted Cumulative Gain for Tourism Recommender System Optimisation," *Proceedings of the 2021 IEEE International Conference on Computing*, pp. 31-36, Nov. 2021. DOI: 10.1109/ICOCO53166.2021.9673502
- [9] G. Jung, and K. Kim, "Development of Machine Learning-Based Predictive Models and a Web App for University Cafeteria Attendance," *Journal of The Korea Society of Computer and Information*, Vol. 29, No. 12, pp. 149-157, Dec. 2024. DOI: 10.9708/jksci.2024.29.12.149
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin, "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*, pp. 1135-1144, Aug. 2016. DOI: 10.1145/2939672.2939728
- [11] S. M. Lundberg, and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Vol. 30, pp. 4765-4774, Dec. 2017. DOI: 10.48550/arXiv.1705.07874

Authors



Yubin So is an undergraduate student in the Department of Management Information Systems at Myongji University, Korea, and is expected to graduate in February 2026. His research interests include machine learning

and data science, and he plans to continue his studies in these fields.



Eunbi Woo is an undergraduate student in the Department of Management Information Systems at Myongji University, Korea, and is expected to graduate in February 2026. She is interested in software development and is

preparing for a career as a developer.



Hanjun Lee received his B.S. and M.S. degrees in Computer Engineering from Seoul National University in 2001 and 2004, respectively, and his Ph.D. degree in Management Information Systems from Korea

University in 2016. He joined the Department of Management Information Systems at Myongji University, Seoul, Korea, as a faculty member in 2020. His research interests include data science, particularly text mining, machine learning, and online review analysis.