

주요 선행연구 비교 검토

	데이터셋(기간)	데이터셋(대상 리그)	데이터셋(선수 수)	모델	평가	연구목적	연구 결과	기타	연구 진행 시기	
우리연구	10 시즌(10-19)	유럽 주요(5개 리그)	약 500명	선형회귀, ML		1. Salary prediction Regression analysis			2020	
Predicting Market Value of Soccer Players Using Linear Modeling Techniques	1시즌(현재인진 모름)	유럽 전체	357명	OLS, KNN, Ridge 회귀, PCR	RMS	1. predicting market value of top players using statistical modeling techniques	1. PCR(k=15)가 제일 좋은 성능			
Football Player's Performance and Market Value	1시즌(14-15)	라리가	[transferred player] - 37명(market value train data : 실제 아직 선수) - 40명 (performance train data : 이 적 시장에 올랐던 40명 데이터 활용)	Lasso regression	K-fold validation	1. prediction for salary and performance 2. relation between salary and performance - how market value and performance of La Liga (the Spanish League) players can be modeled using extensive public data sources. - develop regression models to predict the real market value and assess a player's performance	1. Performance, market 모델 확인 2. Market value와 performance간 관계 확인 (고평가 선수의 overpay는 marketing 비용으로 고려됨)	- Market value proxy : Market value estimate 로 트랜스퍼마켓 자료 활용(proxy) - market value real : transfer fee		
Machine Learning for Soccer Analytics	1시즌 (11-12)	EPL	EPL 선수 전원	ML 알고리즘 다수	MAE(mean absolute error) 등	1. Performance - rating 2. Match outcome - performance 3. Match outcome - rating				
Assessing the market values of soccer players –a robust analysis of data from German 1. and 2.Bundesliga	1시즌 (15-16)	분데스 1,2군	493명	robust regression	평가 X(예측 아님, y-x간 상관관계 분석을 위해 회귀계수 pvalue 값 확인 진행)	1. extent a player's market value depends on his skills	1. performance와 club reputation이 market value에 큰 영향 2. 고/저 평가 선수 존재(marketing 등 영향 고려)	Robust regression 설명 블로그 : https://brunch.co.kr/@gimmesilver/38	2018	
ANALYSIS OF STATISTICS IN MAJOR LEAGUE SOCCER	1시즌 (16)	MLS(미국)	340게임(게임 stasts - 승점 간 관계 연구)	Linear regression(simple, multiple)	- Simple LR = 귀무가설 검정 - multiple LR = 조정 R, BIC로 모델 선택	1. to analyze statistical data collected from all games (n = 340) in the 2016 Major League Soccer regular season and discover which statistics are correlated to total season points.	1. 승점 관련 경기 stats 확인			
Application of Neural and Regression Models in Sports Results Prediction	Javelin 경기 대상	Javelin 경기 대상	Javelin 경기 대상	regression, Perceptron		1. comparing regression and neural models with respect to their accuracy of predicting sports results.	1. the investigation demonstrated a significantly greater accuracy of prediction for perceptron models.		2014	
A Multiple Linear Regression Approach For Estimating the Market Value of Football Players in Forward Position	1시즌 (17-18)	유럽 주요(4개 리그)	105명	OLS regression	MAPE, Rsquared => 회귀 결과 평가, (예측모델 평가 x)	1. to examine if the independent variables are successful in predicting the outcome variable and which independent variables are significant predictors of the outcome	1. Market value의 주요 변수 확인		2018	
Modelling the transfer prices of football player	5개 시즌 (12-17)	유럽 전체	424명	1. Forward stepwise selection 2. lasso 3. Ridgeregression	- 예측 모델 성능 평가 : k-fold - error measure : 1) MeanSquaredError(MSE). 2) RootMeanSquaredError(RMSE). 3) MeanAbsoluteError(MAE). 4) MeanPercentageError(MPE)	1. predicting transfer prices	1. LASSO모델의 예측성능이 제일 좋음 2. 이적료에 영향을 주는 요소들 확인	Squawka Score, European club index(ECI) ECI : https://www.euroclubindex.com/	2017.09	1. [회귀 모델링 참고논문] 회귀분석 모델링의 전체적인 과정이 잘 설명되어있어서, 우리 연구 진행 시, 스탭바 이스텝으로 참고하면 좋을 듯!
Application of Multiple Linear Regression Models in the Identification of Factors Affecting the Results of the Chelsea Football Team	1시즌 (14-15)	첼시팀	38경기 데이터(28개 train, 10개 test, 30명)	1. Regression model(선수 출전시간 - 득점) 2. Regression model(기타 경기 특성)	모델링(p값에 따른 유효한 회귀계수 확인)	1. to identify the influence of a player on the result of a match, by means of the generation of a multiple linear regression model			2017	
Player valuation in European football	1시즌 (15-16)	유럽 주요(5개 리그)	2,606명 (10, 25, 50% 랭크 선수들로 데이터셋 구성, 10% : top / 25% : good / 50% : average)	ML 알고리즘 10개 (RandomForest, BayesNet, Logistic, DecisionTable (a decision table majority classifier), IBk, KStar (nearest neighbor with generalized distance), NaiveBayes, J48, Part, and ZeroR (predicts the majority class for nominals or the average value for numerics)	- 예측 모델 성능 평가 : k-fold - 성능 측정 지표 : accuracy, precision, recall, F1 score (or f-measure), and AUC-ROC.	1. compare and contrast which attributes and skills best predict the success of individual players in their positions in five European top football leagues	1) 개별 선수들의 성공에 영향을 미치는 특징, 스쿼드 확인 2) 다른 연구와 달리, top player(10%, 25%, 50%) 집중 연구 3) 10개 ML 알고리즘의 예측 성능 비교 분석 4) 예측 성능이 포지션 별, 팀 타이틀 자동화(f1 score) * 공격수, 탑 타이틀수록 성공 요인이 더 두드러지게 관찰됨 5) basic ML 기법 : EPL, 분데스리가 선수 예측에 더 높은 성능(basic ML : Zero R classifier)		2018	1. [ML 모델링 참고 논문] 2. [데이터 setting 참고] : LEAGUE-POSITION-Xpc where LEAGUE has the values EPL, Bundesliga, La-Liga, Ligue-1, Serie-A and All; POSITION has the values GK, DF, MD, and FW; and X has the values 10, 25, 50
BEYOND CROWD JUDGMENTS: DATA-DRIVEN ESTIMATION OF MARKET VALUE IN ASSOCIATION FOOTBALL	6시즌 (09-14)	유럽 주요(5개 리그)	4,217명	Regression (actual transfer fee - model) vs (actual transfer fee - crowd source[트랜스퍼마켓]) *model의 train은 crowd source데이터 활용.	- 성능 평가 시, k-fold cross validation 대신, RMSE, MAE 활용 (time-series- based evaluation approach 고민해야 할 듯) - fitness : 회귀 모델의 fitness 보기 위해, AIC, BIC, log-likelihood 값 등 확인	1. estimate players' market values using multilevel regression analysis	1) crowd 예측이 어느정도 유의미함. 선수의 90%(하위 이적료)는 model의 예측이 더 fit하지만, 상위 이적료 10%는 crowd-estimate가 더 fit함을 보임 2) 비 경기력 요소 중 위치, 유투브, 데탕 이 유의미한 상관관계 보임		2017	1. [회귀분석 모델링 참고논문] - 시계열 데이터 회귀분석 평가방법 참고(교차검증?) - 비경기력 요소 포함한 모델링