



탐색적 데이터 분석 및 선형회귀 모델을 활용한 FIFA 데이터 분석

FIFA Data Analyze Using Exploratory Data Analysis and Linear Regression Model

Seungjae Moon

Dept. of IT Transmedia Contents, Hanshin University,
Osan-si, Korea

[Abstract]

This paper uncovered the relationship between abilities and performance by exploratory data analysis(EDA) for FIFA 22 players data and predict overall by creating linear regression model. It provides into which abilities players can improve to achieve higher overall scores by regression models' weights. Additionally, it provides useful information to researchers and gamers who want to accurately understand the abilities required for each position. This will support the development of the gaming sports market and finally expected to be connected to the actual sports industry market.

Key word : Exploratory Data Analysis, Linear Regression, FIFA 22, Sports game, Overall scores, Required abilities

1. 서론

스포츠 시장의 발달과 함께 스포츠 게임 시장 또한 매우 빠른 속도로 성장하고 있다[1]. 대형 게임 제작사 EA의 축구 게임인 FIFA 시리즈는 1993년부터 2023년 현재까지 인기를 얻고 있는 매우 전통 있는 게임이다[2]. 이처럼 스포츠 게임의 역사가 길어지면서, 스포츠 게임의 디테일이 매우 높아졌고, 이는 곧 게임의 전략적 요소로 이어졌다. FIFA 게임은 지속적으로 선수의 몸무게, 키, 체형 등과 같은 요소와 슈팅, 드리블, 패스 능력과 같은 세부 능력을 수치화하여 게임에 반영하고 있다. 이러한 전략적 요소들은 현실 스포츠와 같이 변화에 대한 플레이어들의 빠른 이해와 분석을 요구하게 된다[3].

따라서 본 논문에서는 FIFA 22의 선수 데이터를 기본적인 데이터 분석법과 탐색적 데이터 분석(EDA - Exploratory Data Analysis)을 활용하여 데이터의 분포와 특성을 이해한다. 또한, 분석한 데이터의 상관관계를 이해하고, 선형회귀 모델을 통해 선수들의 오버롤을 예측하고, 오버롤에 영향력을 미치는 변수를 알아본다. 나아가, 포지션별로 요구되는 능력치를 알아본다.

II. 선행 연구

2-1 탐색적 데이터 분석 (EDA-Exploratory Data Analysis)

탐색적 데이터 분석(EDA)은 데이터를 분석하고, 기초통계량에 따른 분석법을 넘어, 데이터 시각화 방법을 사용하여 데이터의 주요 특성을 요약하는 데 사용되는 분석법이다[4]. 이를 통해 데이터의 패턴을 발견하고, 이를 바탕으로 기존의 가설을 수정하거나 새로운 가설을 수립할 수도 있다. 또한, 다양한 각도에서 데이터를 관찰하기 때문에, 오류를 식별하는 데에 도움이 될 뿐만 아니라, 데이터 내의 패턴을 이해하여 이상치 또는 변칙적인 현상을 확인하고, 변수 간 유의미한 관계를 찾는 데 도움이 된다[5]. 탐색적 데이터 분석이 끝나고, 도출된 결과는 기계학습과 같은 정교한 데이터 분석과 모델링에 유용하게 쓰일 수 있다.

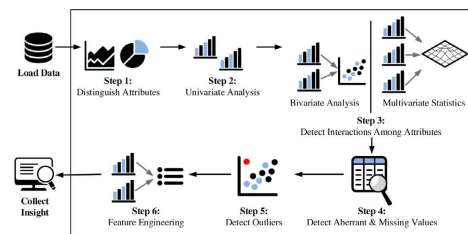


그림 1. 탐색적 데이터 분석 (EDA) [6]

Fig. 1. Exploratory Data Analysis (EDA) [6]

본 논문에서는 탐색적 데이터 분석 기법을 활용하여 FIFA 데이터 세트 내 선수들의 특징 및 세부 능력치 변수 데이터를 분석하여 변수 간 상관관계를 도출한다.

2-2 선형회귀 모델 (Linear Regression Model)

선형회귀란 종속 변수 y 와 한 개 이상의 독립변수 X 의 선형 상관관계를 모델링 하는 회귀분석 기법이다. 관측된 데이터에 선형 방정식을 적용하여 변수 사이의 관계를 선형으로 모델링 하는 것이 주된 목적이다[7]. 이때, 한 개의 독립변수에 기반을 두면 단순 선형회귀, 두 개 이상의 독립변수에 기반을 두면 다중 선형회귀라고 한다[8].

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1)$$

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon \quad (2)$$

기본적으로 단순선형회귀모델은 수식 (1)을, 다중선형회귀 모델은 수식 (2)와 같은 방정식을 따른다.

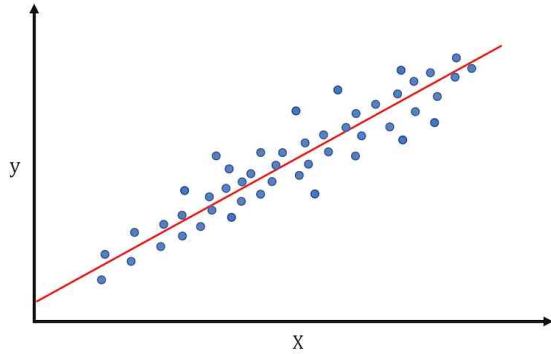


그림 2. 선형회귀 모델 [9]

Fig. 2. Linear regression model [9]

선형회귀 모델은 독립변수와 종속 변수의 관계를 선형 그래프로 나타내고, 해당 그래프와 새로운 독립변수를 통해 값을 예측하는 모델이다.

$$MSE(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_i)\}^2 \quad (3)$$

선형회귀 모델은 독립 변수들의 가중치와 편향을 결정하기 위해 식 (3)과 같은 평균 제곱 오차(Mean Squared Error)라는 손실 함수를 통해 오차를 통해 회귀계수를 평가하고[10], 최소제곱법(Ordinary Least Square)이나 경사 하강법(Gradient Descent) 등을 통해 오차를 최소화한다[11].

본 논문에서는 다중 선형회귀 모델을 이용하여 FIFA 데이터 세트 내 선수들의 오버올 (Overall)을 다양한 독립변수들을 통해 예측하고, 선형성을 확인하여 어떤 변수들이 종속 변수인 오버올에 영향력이 있는지 분석한다.

III. 데이터

3-1 데이터 구성

FIFA에 등록된 선수 데이터를 분석하기 위해 오픈소스 데이터 플랫폼인 Kaggle에서 FIFA22 선수 데이터를 본 연구에 이용한다[12]. 데이터 세트에는 총 19260명의 선수의 'Name', 'Age', 'Overall', 'Nationality', 'BestPosition' 등과 같은 선수 정보 변수와 'Crossing', 'Finishing', 'Acceleration' 등과 같은 능력치 변수 등, 다양한 범주형 변수와 수치형 변수를 합한 총 82개의 변수로 구성되어 있다.

3-2 데이터 전처리

데이터 분석을 시행하기 전 EDA와 선형회귀 모델링에 사용되지 않을 무의미한 변수인 'ID'와 'PhotoUrl' 등의 변수는 데이터 세트에서 제거한다.

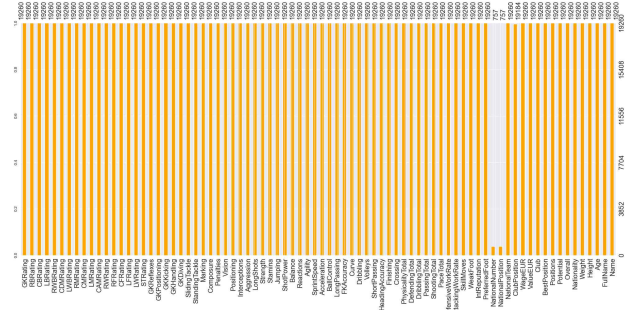


그림 3. 데이터 세트 결측치 그래프

Fig. 3. Dataset's missing values graph

또한, 그림 0과 같이 데이터의 결측치 확인하기 위해 각 변수의 데이터 개수를 시각화한 결과, 'NationalPosition', 'NationalNumber' 두 변수에서 총 18503개, 'ClubPosition' 변수에서 76개의 결측치가 발견되었다. 국가대표로 선발되지 않은 선수들은 NationalPosition과 NationalNumber에 대한 데이터가 존재하지 않다는 것을 알 수 있으며, 결측치 비율이 높아 데이터 세트에서 탈락시킨다. 또한, FA 선수들의 경우 ClubPosition에 대한 데이터가 존재하지 않다는 것을 확인할 수 있으며 해당 결측치는 비율이 높지 않아 새로운 명목형 데이터 'FA'로 보간한다.

IV. 구현 및 테스트 결과

4-1 데이터 분석

데이터에서 유의미한 결과 도출을 위해 데이터 세트 내 여러 가지 변수를 분석한다.

표 1. 국가별 선수 명수

Table 1. Number of Players by Nation

Number of Players by Nation (Total Nation : 163)					
England	Germany	Spain	France	Argentina	...
1718	1214	1019	987	955	...

표 1은 FIFA 22 선수 데이터 내 국가별 선수 명수다. 총 163개의 국가가 존재하며, 잉글랜드 선수가 총 1718명, 독일 선수가 1214명, 스페인 선수가 1019명, 프랑스 선수가 987명으로, 가장 많은 선수를 보유한 상위 5개국 중 4개국이 유럽권 국가임을 알 수 있다.

다음으로 어떠한 선수가 분야별 능력치에서 가장 높은 수치를 보유하는지 알아본다.

표 2. 분야별 최고 능력치 보유선수

Table 2. Players with the highest ability in each field

Aspects	Players
Maximum Potential	K. Mbappé
Maximum Overall Performnace	L. Messi
Best PaceTotal	K. Mbappé
Best ShootingTotal	Cristiano Ronaldo
Best PassingTotal	K. De Bruyne
Best DribblingTotal	L. Messi
...	...
Best GK Kicking	Ederson
Best GK Positioning	S. Handanovič
Best GK Reflexes	J. Oblak

표2와 같이 잠재력과 속도가 가장 높은 선수는 킬리안 음바페, 오버롤이 가장 높은 선수는 리오넬 메시, 슈팅 능력이 가장 높은 선수는 크리스티아누 호날두 등이 있다. 이때, 분야별 최고 능력치를 가진 선수 목록에서 가장 많이 나타난 선수는 리오넬 메시로 총 9가지 분야에서 능력치가 가장 높다. 해당 결과를 통해 이러한 능력치들이 높을수록 오버롤이 높다는 상관관계를 가질 것이라고 유추할 수 있다.

4-2 탐색적 데이터 분석 (EDA) 결과

위와 같은 기본적인 데이터 분석이 끝나고 변수 간 상관관계를 알아보고 분석하기 위해 Seaborn, Matplotlib 등과 같은 시각화 툴을 이용하여 탐색적 데이터 분석을 진행한다.

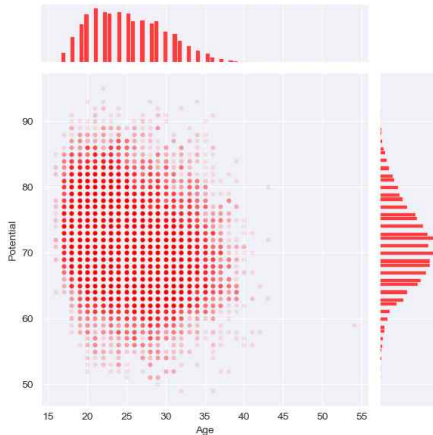


그림 4. 나이와 잠재력 조인트 플롯

Fig. 4. Joint plot for age and potential

그림 0은 나이(Age)와 잠재력(Potential) 두 변수의 분포와 상관관계를 알아보기 위해 나타난 그래프이다. 그래프를 보면 20세에서 30세 사이에 잠재력이 높은 선수들이 많이 분포되어 있으며, 나이가 많아질수록 잠재력이 높은 선수들이 적게 분포한다. 해당 분포를 통해 나이가 많아질수록 잠재력이 떨어진다는 상관관계를 유추할 수 있다.

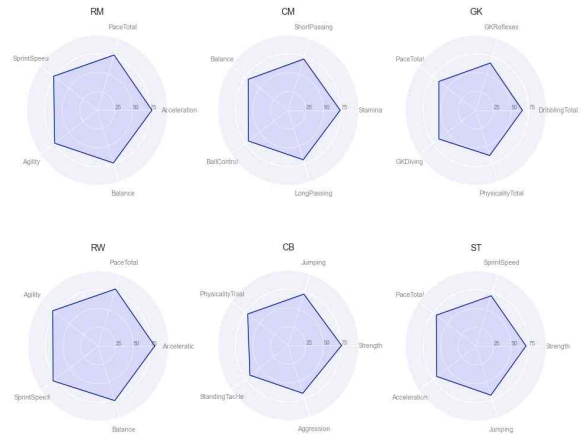


그림 5. 포지션별 가장 높은 평균 능력치

Fig. 5. Highest average stats for each position

그림 0은 포지션별 평균 능력치 중 가장 높은 5가지를 나타낸 레이더 그래프이다. 스트라이커 포지션 선수들의 가장 높은 능력치는 질주 속도, 힘, 점프, 가속도, 슛 파워 능력치였으며, 센터백 포지션 선수들의 가장 높은 능력치는 점프, 힘, 적극성, 스탠딩 태클, 체력이다. 해당 그래프를 통해 포지션별 필요 능력치를 유추할 수 있다.

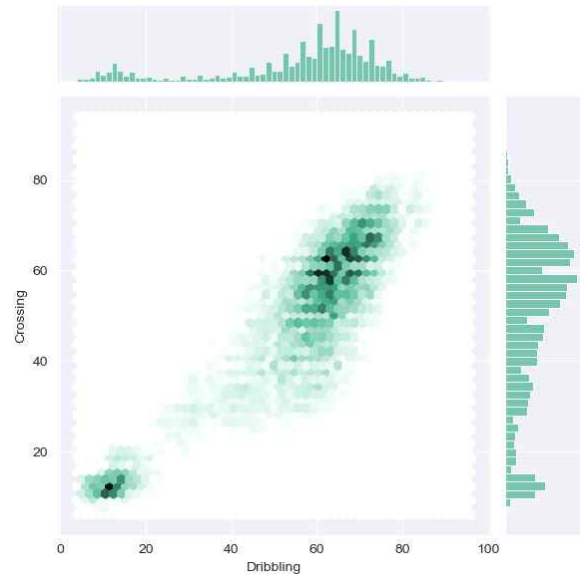


그림 6. 크로스과 드리블 조인트 플롯

Fig. 6. Joint plot for Crossing and dribbling

위 그림은 크로스 능력치와 드리블 능력치의 분포와 상관관계를 알아보기 위한 그래프이다. 해당 분포를 통해 60에서 80 사이의 크로스 능력치와 드리블 능력치를 보유한 선수들이 많은 것을 알 수 있다. 또한, 크로스 능력치와 드리블 능력치가 양의 선형 관계를 띤다는 것을 유추할 수 있다.

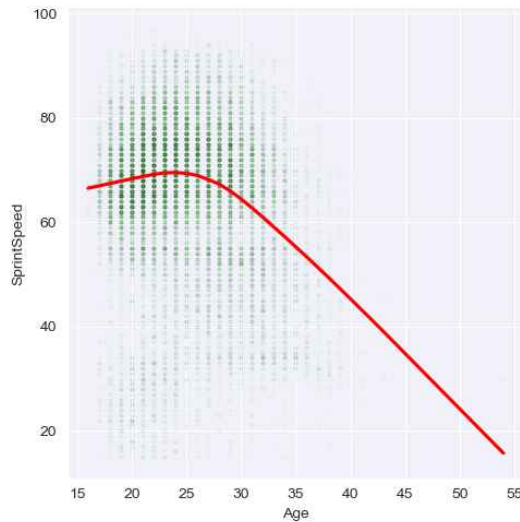


그림 7. 나이와 질주 속도 변수 산점도 및 회귀 그래프
Fig. 7. Scatter plot and regression graph for age and sprint speed

그림 0은 나이(Age)와 질주 속도(SprintSpeed)의 관계를 나타내기 위한 산점도 및 회귀 그래프이다. 20대의 선수층에서 높은 질주 속도를 지닌 선수들이 많이 분포한 것을 확인할 수 있다. 또한, 회귀선을 통해 25세를 기점으로 질주 속도가 절정에 오르며 그 이후로 점점 질주 속도가 느려진다는 것을 유추할 수 있다.

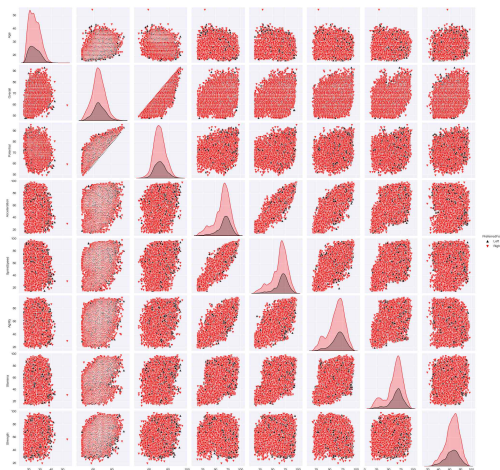


그림 8. 다양한 변수 페어 플롯
Fig. 8. Pair plot of various variables

위 그래프는 주발(PreferredFoot) 변수를 기준으로 나이(Age), 민첩성(Agility), 잠재력(Potential) 등 여러 수치형 변수들과의 상관관계를 표현하기 위한 pairplot 이다. 왼발과 오른발에 따라 검은색과 빨간색으로 다양한 변수를 표시한다. 오른발잡이 선수가 왼발잡이 선수에 비해 많을 것을 확인할 수 있으며, 가속도와 질주 속도가 눈에 띄게 양의 선형의 관계를 이루는 것을 확인할 수 있다. 또한, 민첩성과 질주 속도, 그리고 민첩성과 가속도, 민첩성과 체력 능력치가 비교적 양의 선형 관계를 이루고 있는 것을 확인할 수 있다.

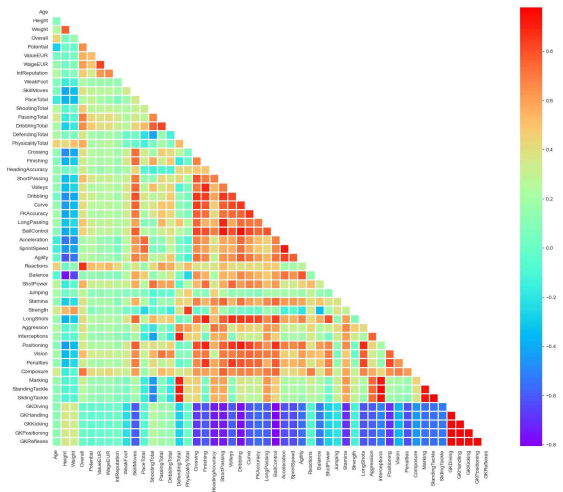


그림 9. 수치형 변수 간 상관계수 히트맵
Fig. 9. Heatmap of correlation coefficients between numeric variables

위 그림은 수치형 변수 간의 선형 상관관계 정도를 나타내는 피어슨 상관계수를 표시한 히트맵이다. 해당 히트맵은 상관계수가 0일 때 상관관계가 없으며, 강한 양의 선형 관계를 가질수록 상관계수가 1에 근접하며 붉은색을, 강한 음의 선형 관계를 가질수록 상관계수가 -1에 근접하며 보라색으로 표시된다. 해당 히트맵을 통해 'Balance' 변수와 'Height' 변수가 강한 음의 상관관계를 가지고, 'Reaction' 변수와 'Overall' 변수가 강한 양의 상관관계를 가진다는 것을 알 수 있으며 다양한 변수 간의 상관 관계를 확인할 수 있다.

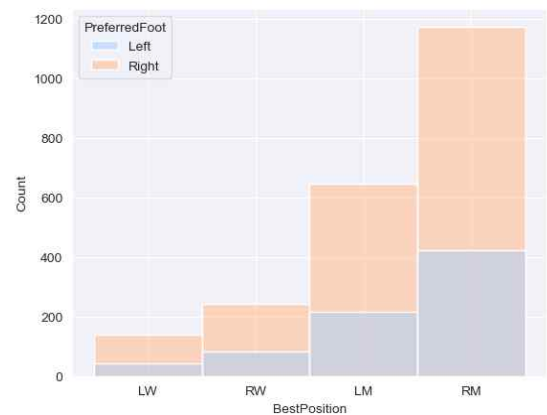


그림 10. 좌·우측 미드필더와 윙어의 주발 히스토그램
Fig. 10. Histogram of LM·RM midfielder and winger's preferred foot

그림 10은 좌·우측 미드필더와 윙어의 주발을 나타낸 그래프이다. 오른쪽 포지션의 경우 오른발잡이가 많았지만, 왼쪽 포지션도 오른발잡이 선수가 많았다. 따라서, 왼쪽 윙어와 미드필더의 경우 역 발 배치 전술을 위한 선수가 많다는 것을 알 수 있다.

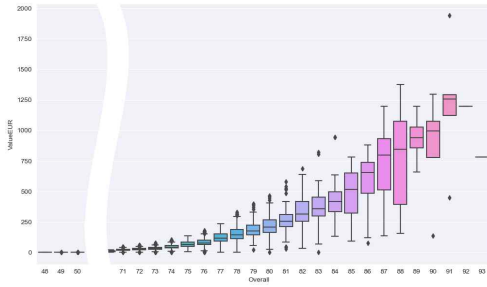


그림 11. 선수 오버롤에 따른 시장가치 박스 플롯

Fig. 11. Box plot of market value by player overall

그림 11은 선수 오버롤에 따른 선수의 시장가치이다. 해당 박스 플롯을 통해 오버롤이 높을수록 선수들의 시장가치 평균이 높고, 시장가치 또한 상대적으로 높은 것을 알 수 있다. 또한, 오버롤이 88인 선수들의 시장가치가 매우 다양한 것을 알 수 있다.

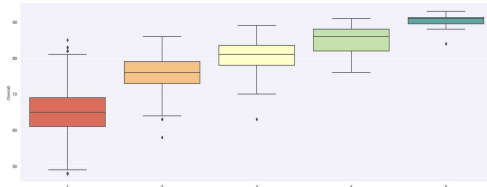


그림 12. 선수 평판에 따른 오버롤 박스플롯

Fig. 12. Overall box plot according to player reputation

그림 12는 선수 평판에 따른 오버롤을 나타낸 그래프이다. 선수들은 1부터 5까지의 평판을 지니고 있으며, 평판이 좋은 선수가 상대적으로 오버롤이 높지만, 선수 평판이 1인 집단을 확인해 보았을 때, 평판이 낮은 선수 중에서도 오버롤이 높은 선수가 존재함을 확인할 수 있다.

4-3 선형회귀 분석 결과

데이터에서 선수들의 오버롤과 다른 변수 간의 선형성을 확인하고 오버롤에 영향력 있는 변수를 찾아낸다. 또한, 포지션별로 어떠한 변수가 오버롤에 영향력 있는지 파악하여 결과를 도출한다. 학습 데이터와 테스트 데이터의 비율은 8:2로 분리하여 진행한다.

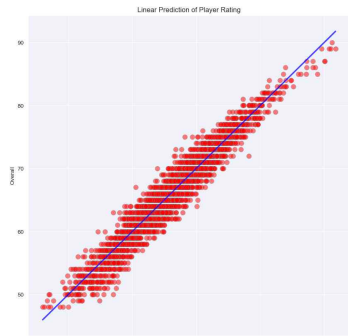


그림 13. 전체 변수 대상 선형회귀 모델 예측 그래프와 가중치 리스트

Fig. 13. Prediction graph of linear regression model and weights list

그림13은 전체 변수를 대상으로 오버롤을 예측하는 선형회귀 모델을 생성하고 예측한 결과를 나타낸 그래프이다. 해당 그림과 같이 데이터가 선형성을 이루고 있으며, 실제값들이 회귀선을 잘 따르는 것을 확인할 수 있다.

표 3. 전체 변수 대상 선형회귀 모델의 r2 score와 RMSE

Table 3. r2 score and RMSE(Root Mean Squared Error) of linear regression for all variables

r2 score	RMSE
0.937	1.669

표 3과 같이 r2 score 값이 0.937로 1에 매우 근접하며, 평균 제곱근 오차(RMSE) 값이 1.669로 매우 작은 값이므로 정확도가 매우 높다고 볼 수 있어 회귀 그래프가 실제 데이터를 잘 설명하며 성능이 높다고 볼 수 있다.

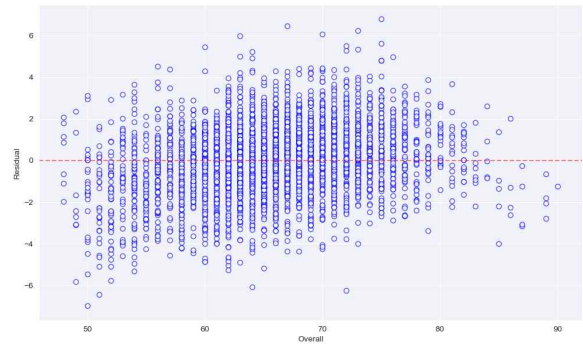


그림 14. 전체 변수 대상 선형회귀 모델의 잔차 산점도

Table 14. Residual scatter plot of linear regression for all variables

그림 14는 해당 선형모델의 유효성 검증을 위해 나타낸 선형회귀 모델의 잔차산점도이다. 몇몇 선수 오버롤의 잔차가 매우 커 이상치로 분류될 수 있지만, 전반적으로 잔차의 분포가 특별한 규칙성을 띠고 있지 않아 해당 선형회귀 모델이 유효하다고 볼 수 있다.

Weight	Feature
0.2641 ± 0.0054	Potential
0.1512 ± 0.0061	Age
0.1135 ± 0.0046	BallControl
0.0944 ± 0.0043	Reactions
0.0409 ± 0.0029	GKReflexes
0.0393 ± 0.0020	ShortPassing
0.0288 ± 0.0014	GKDividing
0.0255 ± 0.0011	HeadingAccuracy
0.0227 ± 0.0009	GKPositioning
0.0218 ± 0.0014	GKHandling
0.0195 ± 0.0008	Stamina
0.0092 ± 0.0014	Dribbling
0.0084 ± 0.0013	Composure
0.0081 ± 0.0011	Crossing
0.0079 ± 0.0006	Finishing
0.0053 ± 0.0010	Acceleration
0.0052 ± 0.0011	Simple_Position_GK
0.0051 ± 0.0007	BestPosition_GK
0.0044 ± 0.0003	SkillMoves
0.0042 ± 0.0003	ValueEUR
...	52 more ...

그림 15. 회귀 모델 변수의 가중치 리스트

Table 15. List of weights of regression model variables

유효성이 검증된 정확도가 높은 모델의 가중치는 각 독립변수가 종속 변수에 얼마만큼의 영향을 미치는지를 나타내는 지표가 될 수 있다. 그림 15와 같이 'Potential', 'Age', 'Ballcontrol', 'Reactions' 변수의 가중치가 상대적으로 높으며, 해당 변수들은 선수의 종속 변수인 오버롤에 상대적으로 영향력이 높다는 것을 확인할 수 있다.

포지션별 요구 능력을 알기 위해 선형회귀 모델을 생성하고 가중치의 크기에 따른 요구 능력을 알아본다. 선형회귀를 생성하기 전, 포지션이 30가지로 너무 많아, ST, AM, CM, DM, DF 총 5가지의 단순 포지션으로 나누어 선형회귀를 진행한다.

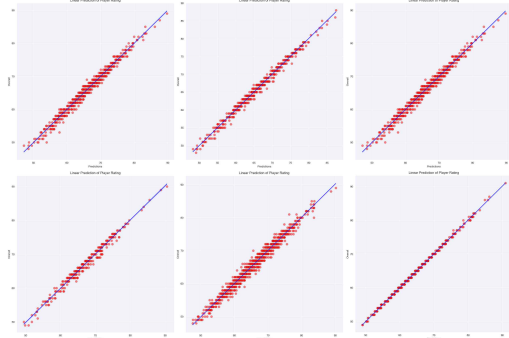


그림 16. 포지션별 선형회귀 모델 예측 그래프

Fig. 16. Prediction graph of linear regression model for each position

위 그림은 포지션별 선형회귀 모델을 생성하고 예측한 결과이다. 해당 그림과 같이 모든 포지션에 대한 선형회귀 모델에서도 데이터가 선형성을 이루고 있으며, 실제값들이 회귀선을 잘 따르는 것을 확인할 수 있다.

표 4. 포지션별 선형회귀 모델 r2 score와 RMSE

Table 4. r2 score and RMSE of linear regression model for each position

	r2 score	RMSE
ST	0.995	0.471
AM	0.989	0.750
CM	0.981	0.944
DM	0.983	0.857
DF	0.978	1.001
GK	0.998	0.324

표 4는 포지션별 선형회귀 모델에 대한 r2 score 값과 RMSE 값이다. r2 score가 1에 매우 근사하며, RMSE 값이 매우 작은 값을 확인함으로써 해당 모델들이 모두 매우 높은 정확도를 나타내고 회귀 성능이 좋다는 것을 확인할 수 있다.

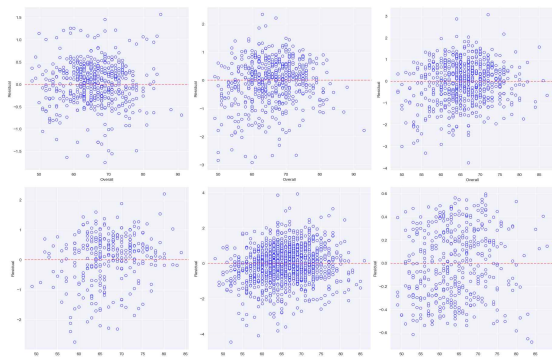


그림 17. 포지션별 선형회귀 모델 잔차 산점도

Fig. 17. Residual scatter plot of linear regression model for each position

해당 선형모델들의 유효성 검증을 위해 그림 17과 같이 잔차 그래프를 확인한 결과, 모든 포지션에 대하여 잔차가 4 이상인 값이 존재하지 않고, 분포 또한 특별한 규칙성을 띠고 있지 않아 선형회귀 모델들이 모두 유효하다고 볼 수 있다.

표 5. 포지션별 오버롤에 미치는 변수의 영향력 순위

Table 5. Ranking of influence of variables on overall by position

	1st Influential	2nd Influential	3rd Influential
ST	Finishing	Positioning	BallControl
AM	BallControl	Dribbling	ShortPassing
CM	BallControl	Dribbling	Potential
DM	Interceptions	Interceptions	Marking
DF	StandingTackle	Marking	Potential
GK	GKPositioning	GKReflexes	GKDividing

표 5는 유효성이 검증된 정확도가 높은 모델의 가중치를 통해 각 포지션에 대한 영향력이 높은 변수 3개를 순서대로 나열한 결과이다. 해당 결과를 통해 공격수 포지션에는 골 결정력, 위치선정, 볼 컨트롤 능력치가, 공격형 미드필더 포지션에는 볼 컨트롤, 드리블, 짧은 패스 능력치가 높을수록 오버롤이 높다는 등, 포지션별 요구 능력치를 도출할 수 있다.

V. 결 론

본 논문에서는 FIFA 22 비디오 게임의 선수 데이터 세트를 기반으로 탐색적 데이터 분석기법을 진행하여 여러 변수 간의 상관관계와 다양한 분석을 진행하였다. 또한, 선형 회귀분석을 통해 선수들의 오버롤을 예측하고, 포지션별로 요구되는 능력치를 도출할 수 있었다. 해당 실험의 다양한 분석을 통해 데이터 분석과 기계학습이 스포츠 게임 전술에 대한 이해를 높일 수 있음을 확인할 수 있었다.

향후 해당 분석법뿐만 아니라 다양한 분석법을 이용하여 선형성을 띠고 있지 않은 종속 변수를 대상으로 비선형 회귀 모델을 생성하여 예측해보거나, 다양한 분류 알고리즘을 통해 여러 변수를 분류해볼 계획이며, 실제 스포츠 산업 분야와 융합되어 능력치에 대한 분석이 연구될수록, 더욱 큰 성장세를 이룰 것으로 기대된다.

References

- [1] LinkedIn, “Sport Games Market Size 2023-2030 Revenue Analysis, Just Updated Research Report”, <https://www.linkedin.com/pulse/sport-games-market-size-2023-2030-revenue>, June. 2023
- [2] Goal, “History of FIFA: Every FIFA game ever released”, <https://www.goal.com/en/lists/history-of-fifa-every-fifa-game-ever-released/eycphoguaggy1fwscnr43azlo>, Feb. 2021
- [3] Jean-Francis Grehaigine, Paul Godbout, “The Foundations of Tactics and Strategy in Team Sports”, *Journal of Teaching in Physical Education*, January. 1999
- [4] Prasad Patil, “What is Exploratory Data Analysis?”, *Medium*, Mar. 2018, <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>
- [5] IBM, “What is exploratory data analysis?”, <https://www.ibm.com/topics/exploratory-data-analysis>, Nov. 2023
- [6] Ghosh et al, “A typical EDA process”, *DEVOPEDIA*, 2018, <https://devopedia.org/exploratory-data-analysis>
- [7] Complete Dissertation, “What is Linear Regression?”, Nov. 2023, <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/>
- [8] Adam Hayes, “Multiple Linear Regression (MLR) Definition, Formula, and Example”, *Investopedia*, April. 2023, <https://www.investopedia.com/terms/m/mlr.asp>
- [9] Martin Tin, “Intro to Linear Regression — Machine Learning 101”, *Medium*. Sep, 2020, <https://medium.datadriveninvestor.com/machine-learning-101-part-1-24835333d38a>
- [10] Jim Frost, “Mean Squared Error (MSE)”, *Statistics By Jim*, 2023, <https://statisticsbyjim.com/regression/mean-squared-error-mse>
- [11] Ganesh Patil, “Understanding & Comparing Linear Regression Using OLS and Gradient Descent”, *Medium*, Aug. 2020, <https://medium.com/@ga3435/understanding-comparing-linear-regression-using-ols-and-gradient-descent-de237c14516>
- [12] ALEX “FIFA 22 Complete Player Dataset [UPD:01/10/21]”, *Kaggle*, Jan. 2021, <https://www.kaggle.com/datasets/cashncarry/fifa-22-complete-player-dataset>