



Football tracking data를 이용한 이벤트 데이터 예측 및 분석

팀명: fm마스터

부산대학교 정보컴퓨터공학부

202055614 최성민

202055502 강동권

지도교수: 이기준

목차

- 1 과제 배경 및 목표
 - 1.1 과제 배경
 - 1.2 과제 목표
- 2 요구 사항 분석
 - 2.1 기존 요구사항
 - 2.2 요구조건 수정사항
 - 2.3 제약사항 분석
- 3 개발 환경 및 사용 기술
 - 3.1 개발 환경
 - 3.2 사용 기술 및 프레임 워크
 - 3.3 외부 리소스 및 참고 데이터
 - 3.4 Workflow Diagram
- 4 갱신된 개발 일정 및 역할 분담
 - 4.1 개발 일정
 - 4.2 역할 분담
- 5 보고 시점까지의 과제 수행 내용 및 중간 결과
- 6 참고문헌

1. 과제 배경 및 목표

1.1 과제 배경

현대 스포츠에서는 데이터 기반의 전략 수립과 경기 분석이 점점 더 중요해지고 있으며, 그 중심에는 트래킹 데이터(tracking data)가 있다. 트래킹 데이터란 경기 중 선수들의 위치, 이동 경로, 속도, 가속도 등 시공간적인 정보를 일정 주기로 수집한 데이터로, 전통적인 이벤트 기반 데이터(예: 패스, 슈팅 기록 등)보다 훨씬 더 정밀하고 방대한 정보를 담고 있다. 이러한 데이터는 선수의 움직임 패턴, 포지셔닝 전략, 팀 전술 분석 등 다양한 목적으로 활용될 수 있다.

특히 축구의 경우 22명의 선수와 공이 넓은 필드 위를 끊임없이 움직이며 복잡한 상호작용을 보이기 때문에, 경기 이해를 위해 트래킹 데이터를 활용한 분석이 필수적으로 여겨지고 있다. 하지만 현재 대부분의 분석 시스템은 사후적이고 정적이며, 실제 경기 상황에서 일어나는 이벤트를 실시간으로 예측하거나 맥락적으로 해석하는 데에는 한계가 있다.

예를 들어, 패스나 슈팅과 같은 이벤트는 단순히 공의 움직임만으로 설명되기 어렵고, 주변 선수들의 위치, 압박 유무, 속도 변화 등 다양한 요소가 결합되어 발생한다. 따라서 트래킹 데이터를 기반으로 한 이벤트 예측 모델은 보다 높은 정확도와 풍부한 경기 해석을 가능하게 하며, 나아가 코칭 스태프나 전술 분석가에게 실질적인 인사이트를 제공할 수 있다.

또한, 머신러닝과 딥러닝 기술의 발전으로 이러한 예측 문제를 해결할 수 있는 가능성이 더욱 커지고 있다. 시계열 분석, 공간적 관계 모델링, 행동 패턴 인식 등의 기법을 트래킹 데이터에 적용하면, 경기 중 발생 가능한 주요 이벤트를 사전에 예측하고, 이를 기반으로 전략적인 결정을 내릴 수 있다.

따라서 본 과제는 축구 트래킹 데이터를 활용하여 이벤트 발생을 예측하고 이를 분석하는 시스템을 개발함으로써, 기존 분석의 한계를 극복하고 보다 지능적인 스포츠 분석 플랫폼의 기틀을 마련하는 데 의의가 있다.

1.2 과제 목표

본 과제는 축구 경기에서 수집된 선수 tracking data를 기반으로 다음의 세 가지 목표를 달성하는 것을 목적으로 한다.

- **이벤트 예측 모델 개발:** 선수들의 위치와 움직임 데이터를 기반으로 패스, 슈팅, 태클 등 주요 이벤트를 예측할 수 있는 모델을 설계 및 구현한다. 여러가지 딥러닝 모델을 Survey하여 모델을 선택하고 Tuning한다.
- **Tracking data 가공:** Football Manager 2024 시뮬레이터(이하 FM2024)를 통해서 축구 경기에서의 Tracking data를 여러가지 기법을 통해 직접 추출하고 가공하여서 데이터 부족 문제를 해결한다.
- **이벤트 분석 및 시각화:** 예측된 이벤트 데이터를 활용하여 경기 흐름, 전략, 선수별 기여도 등을 분석하고 이를 직관적으로 시각화할 수 있는 시스템을 구축한다.

2. 요구 사항 및 제약 사항 분석에 대한 수정 사항

2.1 기존 요구사항

1. FM2024 또는 실제 경기로부터 수집된 트래킹 데이터를 다양한 포맷(CSV, JSON 등)으로 입력 받을 수 있어야 한다.
2. FM2024의 경기 영상을 기반으로 컴퓨터 비전 기술을 통해 각 선수의 좌표 및 이동 경로를 자동 추출하는 기능이 있어야 한다.
3. 추출된 트래킹 데이터를 기반으로 주요 이벤트(패스, 슈팅, 태클 등)를 예측할 수 있는 머신러닝 또는 딥러닝 기반 모델을 사용한다.
4. 예측된 이벤트를 필드 위에 시각화 하고, 분석할 수 있어야 한다.

2.2 요구조건 수정사항

1. 본 과제에서는 iBassek et al. (2025)이 Scientific Data 저널에 공개한 무료 오픈 데이터셋과 Metica Sports에서 공개한 샘플 데이터셋을 사용하였다. 각 데이터셋은 JSON과 CSV 형식으로 구성되어 있고 총 10경기의 공개 데이터셋을 활용할 수 있었고 추가적인 Data augmentation을 활용하였다.
2. FM2024의 경기 영상을 기반으로 좌표는 추출할 수 있었으나, 이벤트 데이터는 자동 추출이 어려워 직접 경기를 진행하며 수기로 기록하는 것으로 바뀌게 되었다.
3. 모델 구조 설계 시에 RNN류(LSTM, GRU) 또는 Transformer 기반 시계열 모델의 성능도 고려해 보면 좋겠다는 피드백을 반영하여 해당 기반 모델로 과제를 수행한다.
4. 기존의 좌표 data만을 넘어서 positional encoding, velocity embedding 등 '위치+속도' 통합 특성 변수를 만들어 보는 것도 예측 성능 개선에 기여할 수 있을 것으로 판단되어 참고한다

2.3 제약사항 분석

1. FM2024에서 경기영상을 통해 좌표데이터를 추출하는 과정에서 선수들의 좌표는 비

교적 정확했다. 반면, 공의 좌표는 이미지를 캡처하는 과정에서 선수들의 좌표와 싱크가 맞지 않는 문제점이 발견되어 해당부분의 정확도를 올리는 것이 주된 과제가 되었다.

2. 이벤트 데이터는 자동 추출할 수 없어 직접 수기로 작성해야 하기 때문에 데이터셋을 만드는 과정의 기간을 늘려야 한다.
3. Yolo모델은 선수 좌표 추출의 정확도가 선수마다 차이가 너무 크고 22명을 동시에 탐지하는 것은 Computing power 및 시간이 많이 소요되어 공의 움직임 만을 추출하는 것으로 목표를 수정해야 한다.
4. 모델 구축에 있어서 RNN 및 Transformer 기반 모델로 기반을 두어 모델 구축을 해야 한다.
5. 모델 평가 및 성능을 측정하는 방법에 있어 기존의 단순한 방식은 모델 평가에 적절치 않다는 피드백을 반영하여 새로운 평가 방식을 찾아야 한다.

3. 개발 환경 및 사용 기술

3.1 개발환경

| 항목 | 내용 |
|----------|---|
| OS | Windows 11 / Ubuntu 22.04 (개인 개발 환경 기준) |
| 프로그래밍 언어 | Python 3.13 / 시각화 툴 추후 수정 |
| 개발 도구 | Jupyter Notebook, Visual Studio Code |
| 버전 관리 | Git, GitHub |
| 가상환경 | conda |

3.2 사용 기술 및 프레임 워크

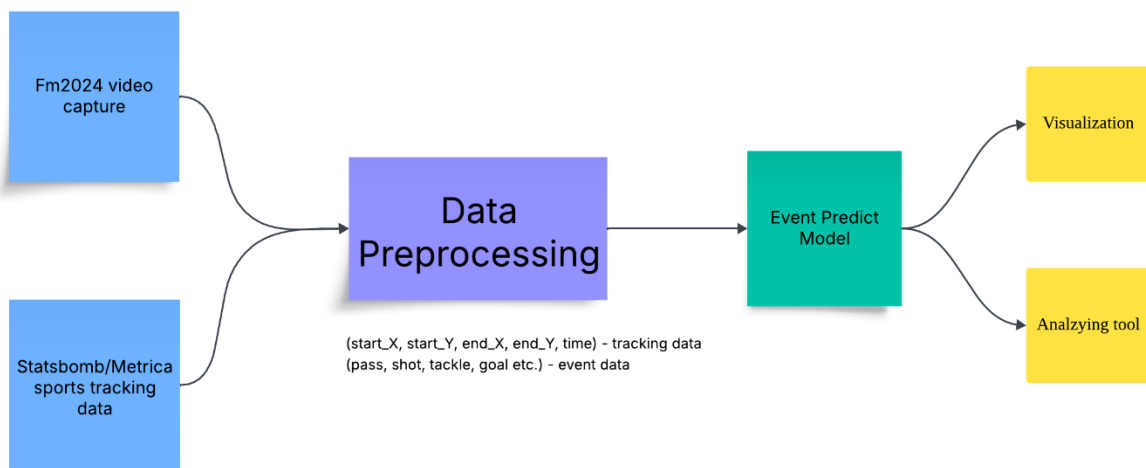
- (1) 영상 처리 및 프레임 추출: FFmpeg, OpenCV

- (2) 객체 검출 및 트래킹: YOLOv8, OpenCV
- (3) 좌표 정규화 및 필드 맵핑: Metrica IO, SciPy
- (4) 이벤트 예측 모델링: scikit-learn, PyTorch/TensorFlow, Optuna
- (5) 데이터 시각화 및 분석: Matplotlib, Seaborn, Plotly, Pandas-Profiling
- (6) 모델 평가 및 성능 측정: Classification Metrics, Cross-validation

3.3 외부 리소스 및 참고 데이터

- (1) Football Manager 2024 video capturing
- (2) 공개 축구 트래킹 데이터:
 - Bassek et al. (2025)이 Scientific Data 저널에 공개한 open 10 games dataset
 - Metrica sports sample 3 games dataset

3.4 Workflow Diagram



4. 갱신된 개발 일정 및 역할 분담

4.1 개발 일정

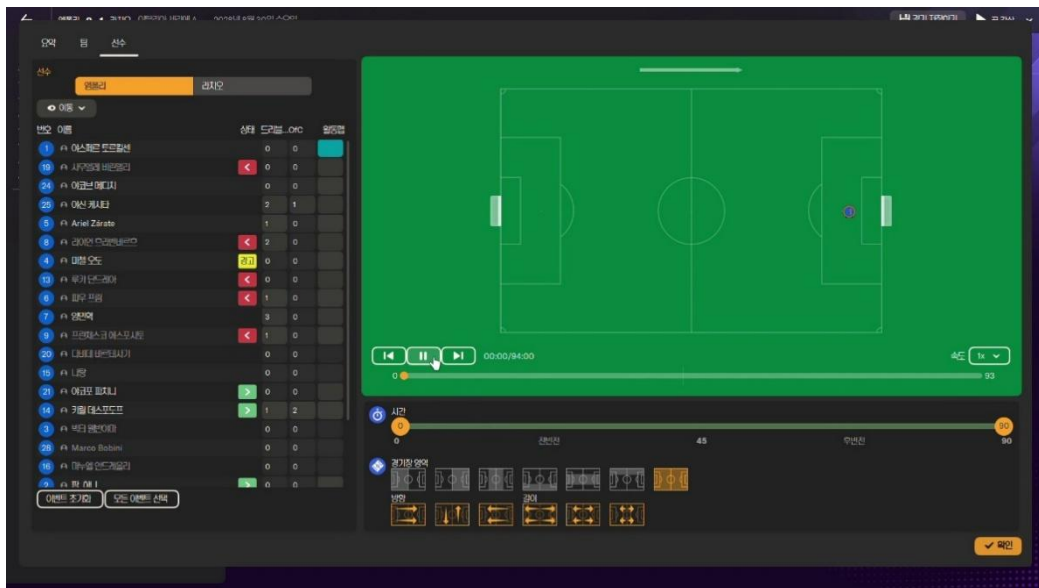
| 구분 | 작업일정 | | | | | | | | | | | | | | | | | |
|----------------------------|------|--|--|----|--|--|----|--|--|----|--|--|----|--|--|--|--|--|
| | 5월 | | | 6월 | | | 7월 | | | 8월 | | | 9월 | | | | | |
| 착수보고서 작성 및 제출 | | | | | | | | | | | | | | | | | | |
| tracking data 처 리 방법 제시 | | | | | | | | | | | | | | | | | | |
| FM2024 데이터 취득 | | | | | | | | | | | | | | | | | | |
| 모델 Survey 및 선택 | | | | | | | | | | | | | | | | | | |
| 중간 보고서 작성 | | | | | | | | | | | | | | | | | | |
| 모델 Tuning 및 최적화 | | | | | | | | | | | | | | | | | | |
| 결과 분석 및 보 수 | | | | | | | | | | | | | | | | | | |
| 시각화 툴 개발 | | | | | | | | | | | | | | | | | | |
| 최종 보고서 작성 및 발표 준비 | | | | | | | | | | | | | | | | | | |

4.2 역할분담

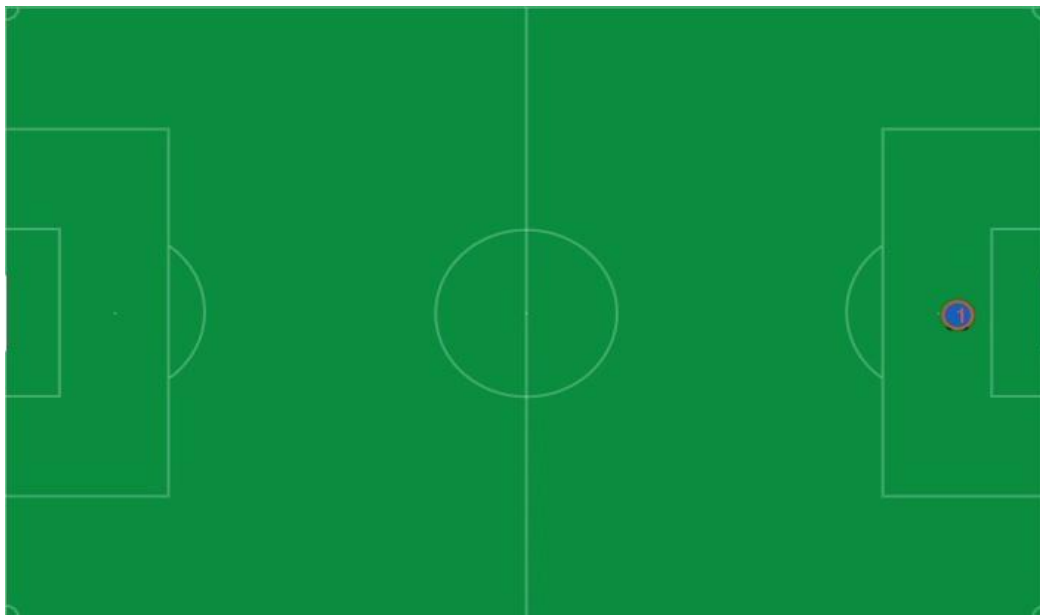
| 이름 | 역할 |
|-----|---|
| 최성민 | FM2024 영상 캡처 및 CSV, JSON화, Event data 작성, 모델 Survey 및 구축, 모델 학습 및 Tuning, 발표 준비, Github page 관리 |
| 강동권 | FM2024 영상 촬영, Ball Trajectory tracking using YOLO, Event data 작성, 모델 Survey, 시각화 툴 개발, 학습 결과 분석 |

5. 보고 시점까지의 과제 수행 내용 및 중간 결과

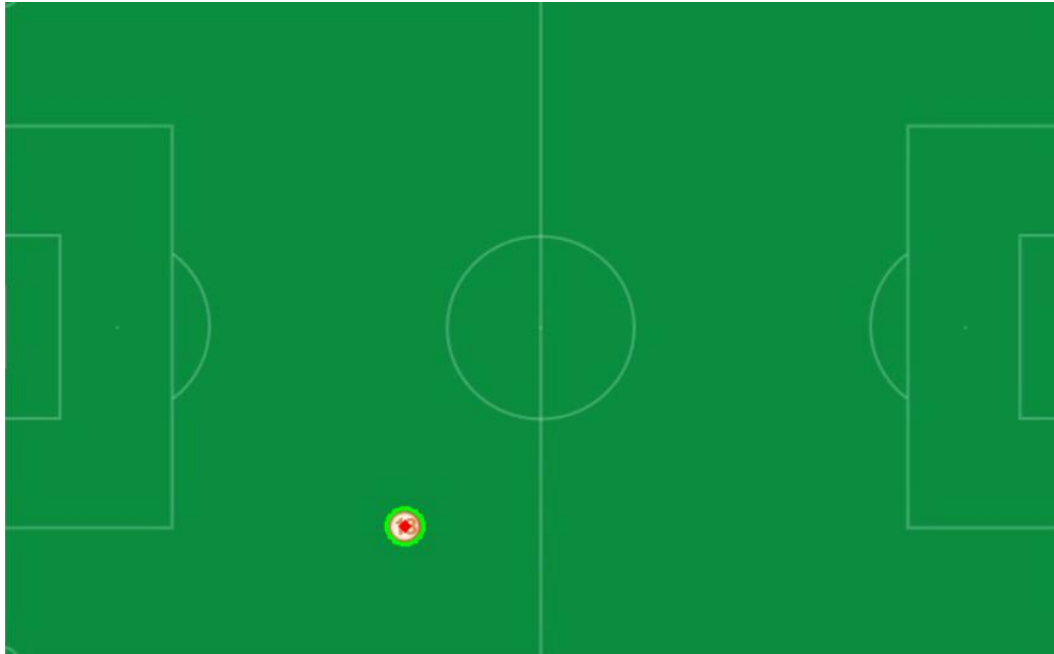
5.1 이미지 샘플링 전체과정



FM2024 비디오를 이미지화 후에 cropping한다.

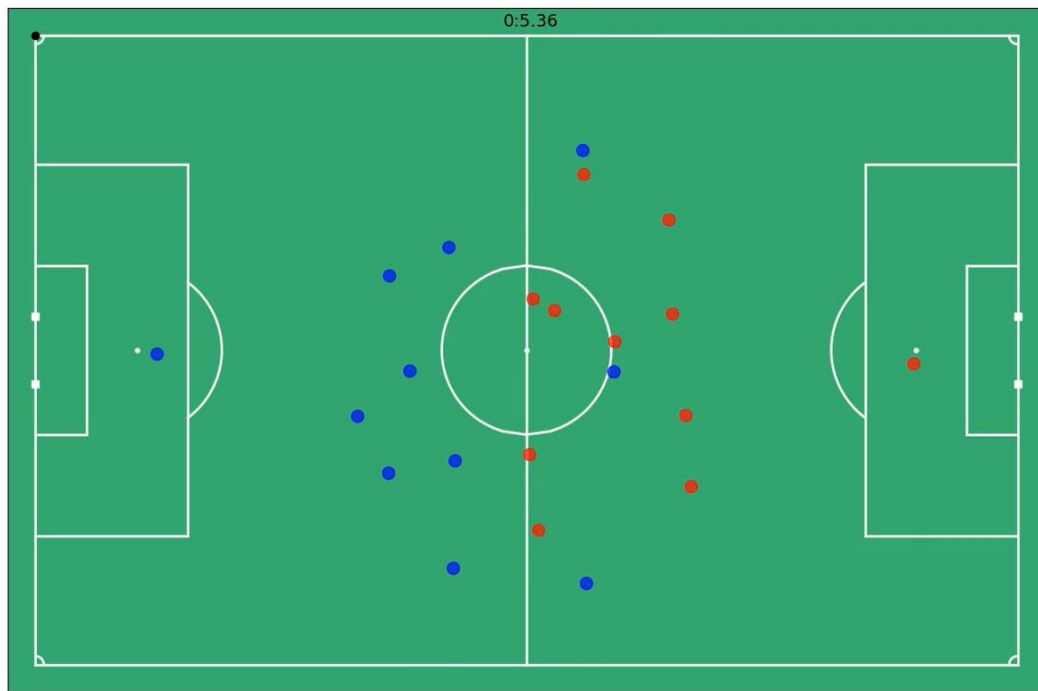


해당 이미지의 선수 위치데이터를 OpenCV의 허프원 변환 등을 이용하여 추출하고 시각화한다.



| Period | Time [s] | Home_1_x | Home_1_y | Home_2_x | Home_2_y | Home_3_x | Home_3_y | Home_4_x | Home_4_y |
|--------|----------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1 | 0 | 43.96590909 | -0.302222222 | 11.74431818 | 0.151111111 | 1.053977273 | 10.27555556 | 4.517045455 | -14.8088888 |
| 1 | 0.04 | 43.96590909 | -0.302222222 | 11.74431818 | 0 | 1.053977273 | 10.27555556 | 4.517045455 | -14.8088888 |
| 1 | 0.08 | 43.96590909 | -0.302222222 | 11.79450758 | 0 | 1.053977273 | 10.22518519 | 4.667613636 | -14.8088888 |
| 1 | 0.12 | 44.11647727 | -0.302222222 | 11.84469697 | 0 | 1.053977273 | 10.17481481 | 4.517045455 | -14.6577777 |
| 1 | 0.16 | 44.26704545 | -0.302222222 | 11.89488636 | 0 | 1.053977273 | 10.12444444 | 4.667613636 | -14.8088888 |
| 1 | 0.2 | 44.41761364 | -0.302222222 | 11.89488636 | 0 | 1.355113636 | 10.12444444 | 4.667613636 | -14.8088888 |
| 1 | 0.24 | 43.81534091 | -0.302222222 | 11.74431818 | 0 | 0.903409091 | 9.973333333 | 4.667613636 | -15.1111111 |
| 1 | 0.28 | 43.81534091 | -0.302222222 | 11.74431818 | 0 | 0.903409091 | 9.973333333 | 4.517045455 | -15.1111111 |
| 1 | 0.32 | 43.890625 | -0.226666667 | 11.74431818 | 0 | 0.903409091 | 9.973333333 | 4.441761364 | -15.1111111 |
| 1 | 0.36 | 43.96590909 | -0.151111111 | 11.74431818 | 0 | 0.903409091 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.4 | 43.890625 | -0.226666667 | 11.74431818 | 0 | 1.016335227 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.44 | 43.81534091 | -0.302222222 | 11.81960227 | 0.151111111 | 1.129261364 | 9.973333333 | 4.517045455 | -15.1111111 |
| 1 | 0.48 | 43.96590909 | -0.151111111 | 11.89488636 | 0.302222222 | 1.2421875 | 9.973333333 | 4.517045455 | -15.1111111 |
| 1 | 0.52 | 43.96590909 | -0.151111111 | 11.84469697 | 0.251851852 | 1.355113636 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.56 | 43.94917929 | -0.167901235 | 11.79450758 | 0.201481481 | 1.053977273 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.6 | 43.93244949 | -0.184691358 | 11.74431818 | 0.151111111 | 1.154356061 | 9.973333333 | 4.441761364 | -15.1111111 |
| 1 | 0.64 | 43.9157197 | -0.201481481 | 11.74431818 | 0.151111111 | 1.254734848 | 9.973333333 | 4.517045455 | -15.1111111 |
| 1 | 0.68 | 43.8989899 | -0.218271605 | 11.74431818 | 0.151111111 | 1.355113636 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.72 | 43.8822601 | -0.235061728 | 11.74431818 | 0.151111111 | 1.053977273 | 9.973333333 | 4.366477273 | -15.1111111 |
| 1 | 0.76 | 43.8655303 | -0.251851852 | 11.59375 | 0.226666667 | 1.053977273 | 9.973333333 | 4.441761364 | -15.1111111 |
| 1 | 0.8 | 43.84880051 | -0.268641975 | 11.44318182 | 0.302222222 | 1.204545455 | 9.973333333 | 4.517045455 | -15.1111111 |

한 경기에서 Home Player와 Away Player로 나누어 데이터를 수집한다.



이벤트 데이터의 형식은 다음과 같다.

| Team | Type | Subtype | Period | Start Frame | Start Time | End Frame | End Time | From | To | Start X | Start Y | End X | End Y |
|------|-----------|-----------|--------|-------------|------------|-----------|----------|----------|----------|---------|---------|-------|-------|
| Away | SET PIECE | KICK OFF | 1 | 1 | 0.04 | 0 | 0 | Player19 | | NaN | NaN | NaN | NaN |
| Away | PASS | | 1 | 1 | 0.04 | 3 | 0.12 | Player19 | Player21 | 0.45 | 0.39 | 0.55 | 0.43 |
| Away | PASS | | 1 | 3 | 0.12 | 17 | 0.68 | Player21 | Player15 | 0.55 | 0.43 | 0.58 | 0.21 |
| Away | PASS | | 1 | 45 | 1.8 | 61 | 2.44 | Player15 | Player19 | 0.55 | 0.19 | 0.45 | 0.31 |
| Away | PASS | | 1 | 77 | 3.08 | 96 | 3.84 | Player19 | Player21 | 0.45 | 0.32 | 0.49 | 0.47 |
| Away | PASS | | 1 | 191 | 7.64 | 217 | 8.68 | Player21 | Player22 | 0.4 | 0.73 | 0.32 | 0.98 |
| Away | PASS | | 1 | 279 | 11.16 | 303 | 12.12 | Player22 | Player17 | 0.39 | 0.96 | 0.49 | 0.98 |
| Away | BALL LOST | INTERCEPT | 1 | 346 | 13.84 | 380 | 15.2 | Player17 | | 0.51 | 0.97 | 0.27 | 0.75 |
| Home | RECOVERY | INTERCEPT | 1 | 378 | 15.12 | 378 | 15.12 | Player2 | | 0.27 | 0.78 | NaN | NaN |
| Home | BALL LOST | INTERCEPT | 1 | 378 | 15.12 | 452 | 18.08 | Player2 | | 0.27 | 0.78 | 0.59 | 0.64 |
| Away | RECOVERY | INTERCEPT | 1 | 453 | 18.12 | 453 | 18.12 | Player16 | | 0.57 | 0.67 | NaN | NaN |
| Away | BALL LOST | HEAD-INTI | 1 | 453 | 18.12 | 497 | 19.88 | Player16 | | 0.57 | 0.67 | 0.33 | 0.65 |
| Away | CHALLENGE | AERIAL-LO | 1 | 497 | 19.88 | 497 | 19.88 | Player18 | | 0.38 | 0.67 | NaN | NaN |
| Home | CHALLENGE | AERIAL-WI | 1 | 498 | 19.92 | 498 | 19.92 | Player2 | | 0.36 | 0.67 | NaN | NaN |
| Home | RECOVERY | INTERCEPT | 1 | 498 | 19.92 | 498 | 19.92 | Player2 | | 0.36 | 0.67 | NaN | NaN |
| Home | PASS | HEAD | 1 | 498 | 19.92 | 536 | 21.44 | Player2 | Player9 | 0.36 | 0.67 | 0.53 | 0.59 |
| Home | PASS | | 1 | 536 | 21.44 | 556 | 22.24 | Player9 | Player10 | 0.53 | 0.59 | 0.5 | 0.65 |
| Home | BALL LOST | INTERCEPT | 1 | 572 | 22.88 | 616 | 24.64 | Player10 | | 0.5 | 0.65 | 0.67 | 0.44 |

졸업과제 수행 깃허브 레포지토리 주소

- https://github.com/Csenongmin/fm_Master
- Spatio-Temporal Databases Laboratory Private 레포지토리

6. 참고문헌

- ⁱ Bassek, M., Rein, R., Weber, H., Memmert, D. (2025). An integrated dataset of spatiotemporal and event data in elite soccer. Scientific Data, 12(1), 195. <https://doi.org/10.1038/s41597-025-04505-y>