**Project Week 2**

Team MSD-07

Westcliff University

DATA 200

Professor Regmi

May 22, 2025

**Literature Review**

**Article 1**

Title: Forecasting binary outcomes in soccer

Authors: (2021) Annals of Operations Research

Link: <https://link.springer.com/article/10.1007/s10479-021-04224-8>

Review:

This study proposes a statistical model framework for predicting binary events in football, such as win/loss, in top-tier leagues like the English Premier League and Serie A. Using score-driven models essentially logistic regressions tailored to event-specific outcomes they evaluated matches by analyzing binary outcomes (e.g., team wins, red cards). Their logistic models achieved strong predictive performance in both leagues.

Demonstrates a solid data-driven framework for binary outcome prediction in football. Highlights the adaptability of logistic regression to different match outcomes (win/loss, over/under). Encourages focusing on specific binary variables beyond simple wins, expanding your model's scope.

**Article 2**

Title: Factors Associated with Match Result and Number of Goals Scored and Conceded in the English Premier League Authors: Günal Bilek & Betül Aygün (2022)

Link: <https://dergipark.org.tr/en/pub/bitlisfen/issue/68527/1015215>

Review:

The authors investigated 17 performance indicators (e.g., shots on target, possession, clearances) plus situational variables like home advantage and scoring first. They used multinomial logistic regression to model win/draw/lose probabilities and Poisson regression for goals. Results showed that scoring first and match location heavily influenced match results, and variables like shots and clearances significantly impacted goal differentials.

Validates the inclusion of features like “scoring first,” “shots on target,” and “home advantage.” Supports using multinomial logistic regression if you expand beyond binary classification. Enhances your feature engineering by incorporating nuanced in-game statistics.

**Article 3**

Title: A Comparative Study of Logistic Models Using an Asymmetric Link: Modelling the Away Victories in Football

Authors: (2018) Symmetry (MDPI)

Link: <https://www.mdpi.com/2073-8994/10/6/224>

Review:

This paper tackles the skewed distribution of results in football—specifically, the fewer away victories compared to home wins or draws. It compares classic logistic regression and skewed (asymmetric) logit models using European league data. The skewed logit models performed better under class imbalance, improving predictions of away wins.

Addresses class imbalance in match outcomes (especially fewer away wins), a common issue for logistic models. Encourages you to consider asymmetric/logit link functions if imbalance affects your data. Provides advanced modeling techniques to improve away-win predictions.