Kan-Anek Tantitayapong 6380612

Section 1: Topic and problem statement

Project title: “Predicting EURO 2024”

We want to build a predictive model that predicts the outcomes of the EURO 2024 tournament for each game from the group stage down to the final. The ultimate question is who will win the tournament.

Section 2: Introduction and relevant literature

Many models have been built to predict competitions that follow a tournament bracket structure. This boils down to first predicting the outcome of each match in the group stages. This is sensitive to error since a wrong prediction will influence the position of the teams’ positions in the group. We aim to build this model using historical data and statistics and running multiple simulations.

Section 3: Describe your research data

Where is the source of data collection?

The data we’ll be using is a completed dataset from Kaggle.

What is the data about?

The main dataset includes information and statistics of every football game played at an international stage (one country vs another country), excluding the Olympics, from 1872 to 2024. Furthermore, there is a penalty shootout dataset in the case that the match ends in a draw and a tiebreaker is needed. A sub-dataset we will also use is the current FIFA rankings of each team. This will help us determine the current strength and skill level of each team with respect to the others.

Define variables/features in the data you are going to use to perform the analysis/build the model

The features are the venue, tournament, the location of the match, the two teams and the score of the match. We will use this as a training set to try to predict the winner of a new, unseen match, given all the details of the fixture.

Section 4: Exploratory data analysis

Submitted separately as a .ipynb file.

Section 5: Research design

Determine factors that lead to the success of teams in EURO competitions (historically), then have those factors be part of the final model.