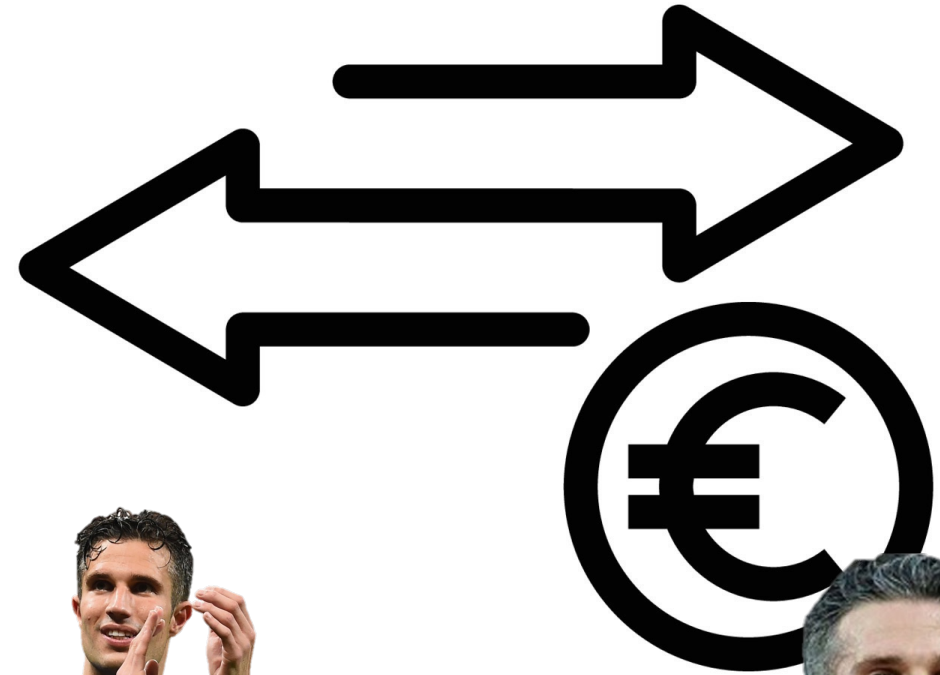


Project Proposal

Unsupervised NER Football Transfer News

Group KFP

- Hugo Lysdahl Hoydal – huly@itu.dk
- Sander Engel Thilo - saet@itu.dk
- Christian Sneftrup Fleischer – chfl@itu.dk



Overall Idea

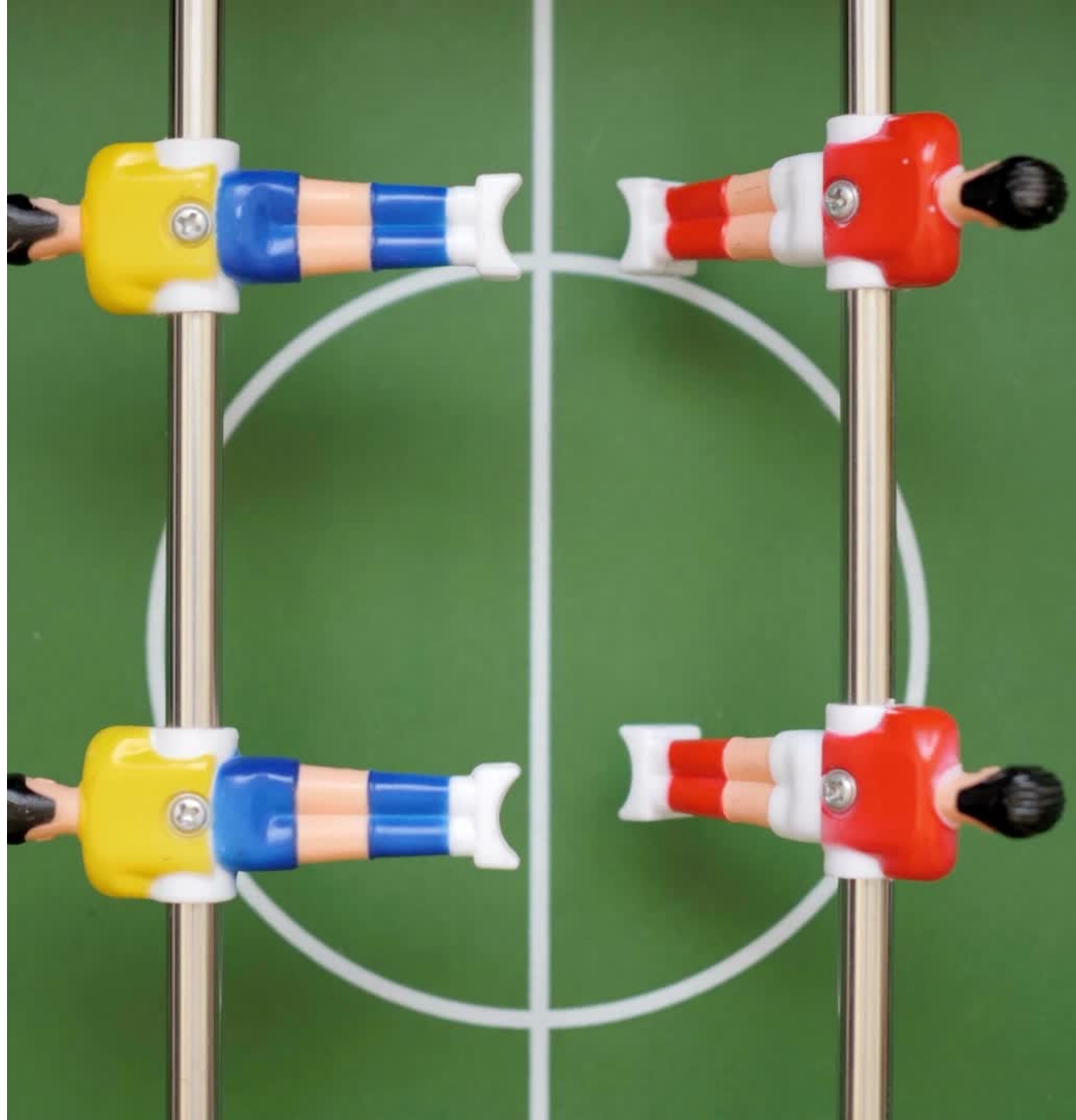
- Unsupervised Named Entity Recognition (NER)
 - Identifying different Named Entities in articles
 - Ex. [player_name], [club_name], [transfer_fee] etc.
- Dataset from 90min.com [1]
 - Transfer News
 - *Title, Date, URL, Content/Text*
- Current status on this topic?
 - NER commonly used for tagging named entities in various ways
 - Unsupervised learning drops the requirement of data labelling
 - No work done regarding football transfers as of yet (from our knowledge) – new part

[1] <https://www.kaggle.com/datasets/crxxom/football-transfer-news-for-nlp>



Baseline Model

- Simple-rule based approach
 - Tag entities based on a predetermined set of rules (eg. Capitalization)
- Advanced Model
 - Pre-trained transformer-based model (BERT or variant)
 - Adapt/fine-tune to football data



Work-in-Progress Research Question

- *How can we apply unsupervised NER models to effectively extract named entities from unlabelled football transfer news articles?*

