Literature review:

|  |  |  |
| --- | --- | --- |
| Article | Features | Model used and accuracy |
| **Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer**  **Youri Geurkink 1**  **, Jan Boone 1,2 , Steven Verstockt 3,† and Jan G. Bourgois 1,2,\*,†** | * Total shots on target in penalty box * Number of accelerations > 2m/s2 second half * Bench current estimated total transfer value * Number of runs > 15km/hr second half * ELO rating * # of direct play possession phases * # of counter attack possession phases * Home/Away * Number of possession won * Distance > 25km/hr second and first half * Distance 6.15km/hr second half * # of crosses | * Variance inflation factor (threshold 5) was used to avoid multicollinearity. * BorutaShap used for feature selection. * Extreme Gradient Boosting for Borutashap and predictiong game outcome. * TreeExplainer to identify strongest predictors. * Accuracy 89.6, precision 88.9, recall 90.1 |
| **Exploring and modelling team performances of the Kaggle European Soccer database**  **Maurizio Carpita1, Enrico Ciavolino2 and Paola Pasca2** | * Forward, Def, Mid, Att * Def and mid important | * Binomial Logistic Regression model * Better at predicting wins compared to RF, NN, k-NN |
| Prediction models for soccer sports analytics  Edward Nsolo  Supervisor : Niklas Carlson Examiner : Patrick Lambrix | Wrapper and filter selection method fop | * Wrapper and filter selection method for feature selection. * 7 ML algorithms to elavuate subsets. Best-first search algorithm and backtrack 5 consecutive non-improving nodes to terminate. |
| **Predicting Wins, Losses and Attributes’ Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis** | * Total Team Medium Pass Attempted * Total Team delivery into the attacking third * Distance Covered Team Average all Zones * Team Average Sprints * Team Average Covered Distance Not in Ball Possession * Total Team clearances attempted * Total Team clearances completed * Ball Possession * Ball Possession Heat Map Attack Field Middle * Total Team tackles gaining the ball * Team Average solo runs into the attacking third * In attacking third * Distance Covered Team Total Zone 4 * Total Team Long Pass Attempted * Team Average clearances completed * Total Team Long Pass Completed * Team Average tackles not gaining the ball * Total Team Long Pass Completed * Total Team Medium Pass Attempted Total Team Short Pass Completed Total Team Short Pass Attempted Total Team Pass Completed Total Team Pass Attempted | * NN with RBF |
|  |  |  |
|  |  |  |
|  |  |  |