# Literature Review Source Notes

## “Outcome prediction of DOTA2 using machine learning methods”

* 3. Data Sets - Data Set II -> all winning teams had at least twice the score of their opponents.
* Naïve Bayes model greatly simplifies learning by assuming that features are independent from one another.
* 6. Experiment and Analysis – scikit learn was used
* “with the increase of the number of features, the prediction accuracy also increases”.
* The paper intrigues the idea of examining individual players.

## Prediction of Football Match Results Based on Model

## Fusion

* “existing prediction models of football matches have problems of poor generalization ability and low accuracy”.
* Proposes the use of model fusion
* Obtained the data of the Chinese super league from 2013-2018.
* Features are selected from the winning features of a football match.
* Three machine learning methods which are Support

Vector Machine (SVM), random forest and Bayesian, are used as a

primary classifier to separately train the data.

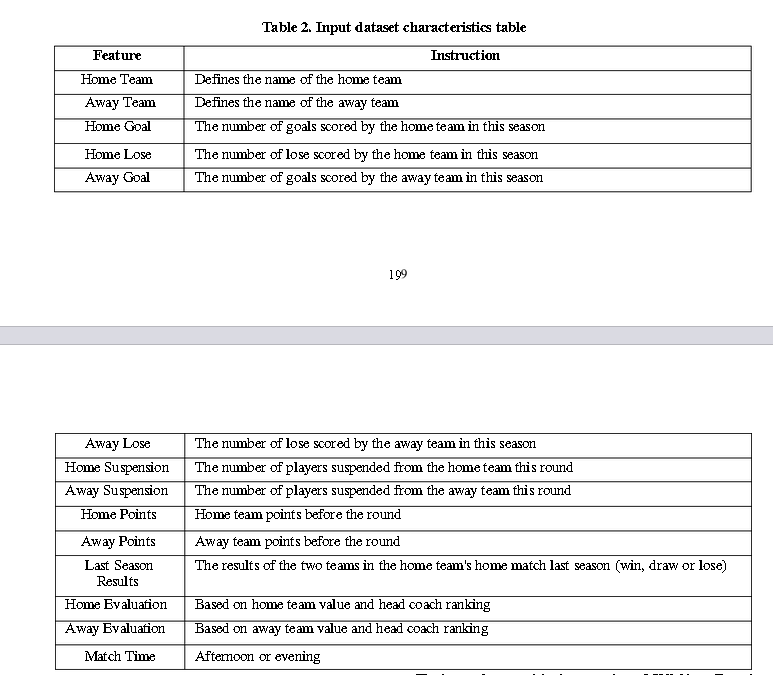
* At last, SoftMax is used as a secondary classifier to establish a prediction model. By

comparing the training results of a single classifier, our method

can achieve better prediction accuracy.

* Can help teams “adjust to a certain extent before the

game and deploy more effective tactics.

* Too many uncertain factors (e.g home/away, weather, game time and referee scale etc..).
* Forecasting methods mainly included the Bayesian model, however this model is too focused on the historical victory or defeat.
* “The three selected single models have good diversity, less correlation, and similar performance, which meet the basic conditions of model fusion.”
*  SVM(support vector machine) is used because it shows a big advantage when working with small samples. Using seasons from too long ago is “not ideal”. Therefore used SVM with less overall training data.
* Random Forest regression model – can process high-dimensional data without feature selection. It can also give important characteristics after training.
* Conclusion – model fusion more successful than single classifier. Better at generalising.

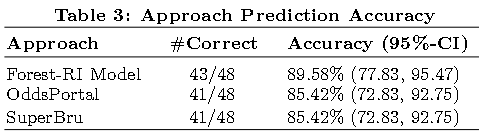
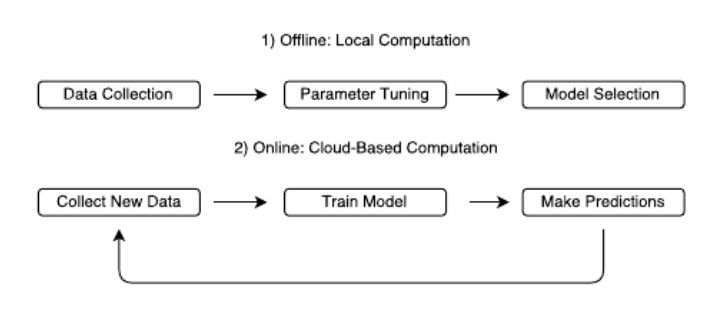
## Human Decision Making and Artificial Intelligence: A Comparison in the Domain of Sports Prediction

* Random forest classifier was used.
* Purpose to compare the predictive ability of a human agent to that of a machine learning prediction model.
* Kahneman and Tversky note that three

classes of information are required when making predictions:

background information, speciﬁc contextual information and

information about the expected accuracy of the prediction.

* Bayes classifier, no matter how optimal will rarely achieve perfect classification due to the intrinsic probabilistic nature underlying observable systems.
* Collected past data of rugby matches
* Trained and compared different forests
* Cloud system used to collect the most recent match data after every match.
* Retrained using the newly incorporated data.
* Compared with aggregate results by human agents “SuperBru”.
* Arguably the most popular random forest algorithm for

classiﬁcation is Breiman’s Forest-RI

## Football Match Result Prediction Using the Random Forest Classifier

* Used 3 previous seasons of football data acquired from [www.premierleague.com](http://www.premierleague.com)
* “The experimental results found that the efficiency of football competition results prediction is good in the accuracy and the precision rates, while the recall rates are acceptable.
* “The classification model for the football match prediction is expected to perform better when there are three types of match results (the home team win, the away team win, and the draw) in the same ration for running the training data”

## Exploring polynomial classifier to predict match results in football championships | Expert Systems with Applications: An International Journal

* “We present an approach to identify the winning team based on the polynomial classifier.”
* “The investigated groups were different types of combinations of two by two pairs, win-draw, win – defeat and draw – defeat.”
* “The proposed system applies a polynomial algorithm to analyse and define matches results.”
* “The association between polynomial algorithm and machine learning techniques allowed a significant increase of the accuracy values. Our polynomial algorithm provided an accuracy superior to 96%, selecting the relevant features from the training and testing set.”

## A Machine Learning Framework for Sport Result Prediction

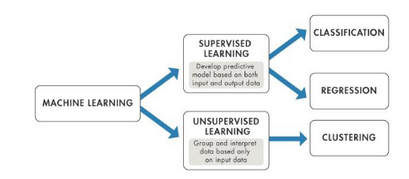
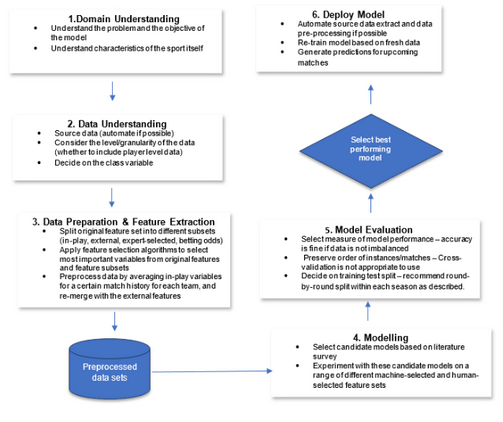
* “…numerous factors involved in the games, such as the results of historical matches, player performance indicators, and opposition information.”
* Focuses on the application of Artificial Neural Networks.
* “Artificial Neural Networks (ANNs) (Grossberg, 1988) are perhaps the most commonly applied approach among ML mechanisms to the sport result prediction problem. “
* 

Figure - Steps of our proposed SRP - CRISP -DM framework. ( Cross-industry standard process for data mining.