Lets Kick Off !!!!

PREDICTING FOOTBALL MATCH WINNER

KUSHAL GEVARIA 60004120034
HARSHAL SANGHAVI 60004120085
SAURABH VAIDYA 60004120115

Guided by:Khushali Deulkar

INTRODUCTION

- Features influence the results of the matches
- Select most significant parameters
- Machine learning and data mining used

PROBLEMS FACED

- Difficulty in predicting draw class
- Lack of data sources
- High entropy : The distribution of wins,losses and draws was 35.5%, 35.5% and 29% respectively. So if we calculate the measure of randomness :

  Entropy = -(0.29 * log3(0.29)) + 2(0.355 * log3(0.355)))

  =0.72

PREVIOUS SYSTEMS

- Earlier method used past result of 10 seasons, form etc.
- Drawback: 10 seasons very long and inaccurate.
- Other systems used various other features.
- Drawback: differences in computation and parameters.

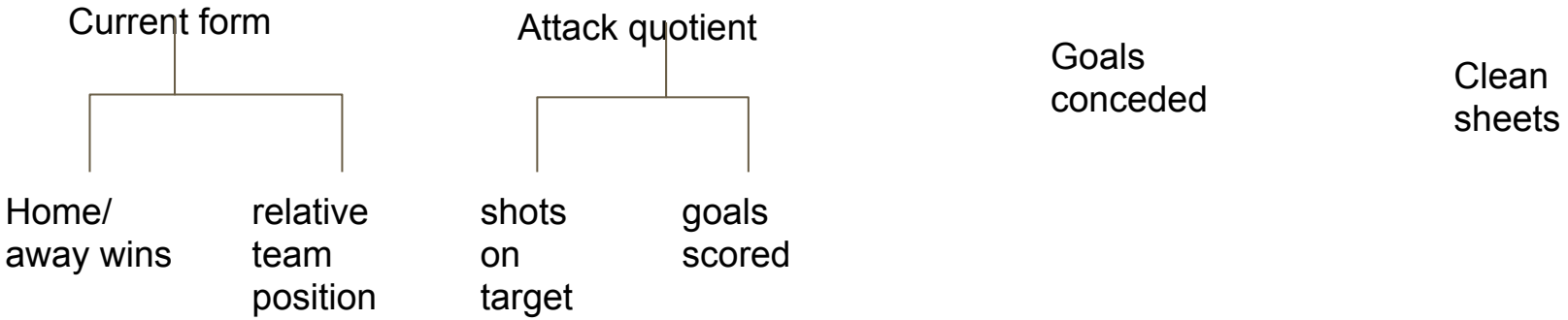
COMPARATIVE STUDY

Author	Classifiers	Features	Accuracy
Douwe Buursma	Classification via regression	Goals scored, Goals conceded, Average points per match, Number of wins (home and away)	After testing of all classifiers, the best average accuracy obtained was 55.08% using Classification via regression and multi-class classifier
	Multi-class classifier		
	Rotation forest		
	Logit boot		
	BayesNet		
	Naïve Bayes		
	Home wins		

Author	Classifiers	Features	Accuracy
Nivard van Wijk	Random probability and team grouping.	Frequency of number of goals per match, Average number of points per match	48%
	Multi independent model		47.24%
	Single independent model		53.03%
	Dependent model		53.55%
Ben Ulmer	Baseline	Results of the matches of past 10 years, current form of 7 matches	40%
	Naïve base		48%-44%
	Hidden markov model		48%-44%
	SVM		55%
	Random forest		49%

IMPLEMENTATION : FEATURES

- The prediction system implemented by us has 4 main parameter components:



FEATURES : FORM

Teams	Points	Multiplying factor	Home Loss	Away Win
A	0.75	0.15	-20%	20%
B	0.6	0.4	-16%	16%
C	0.4	0.6	-12%	12%
D	0.15	0.75	-10%	10%

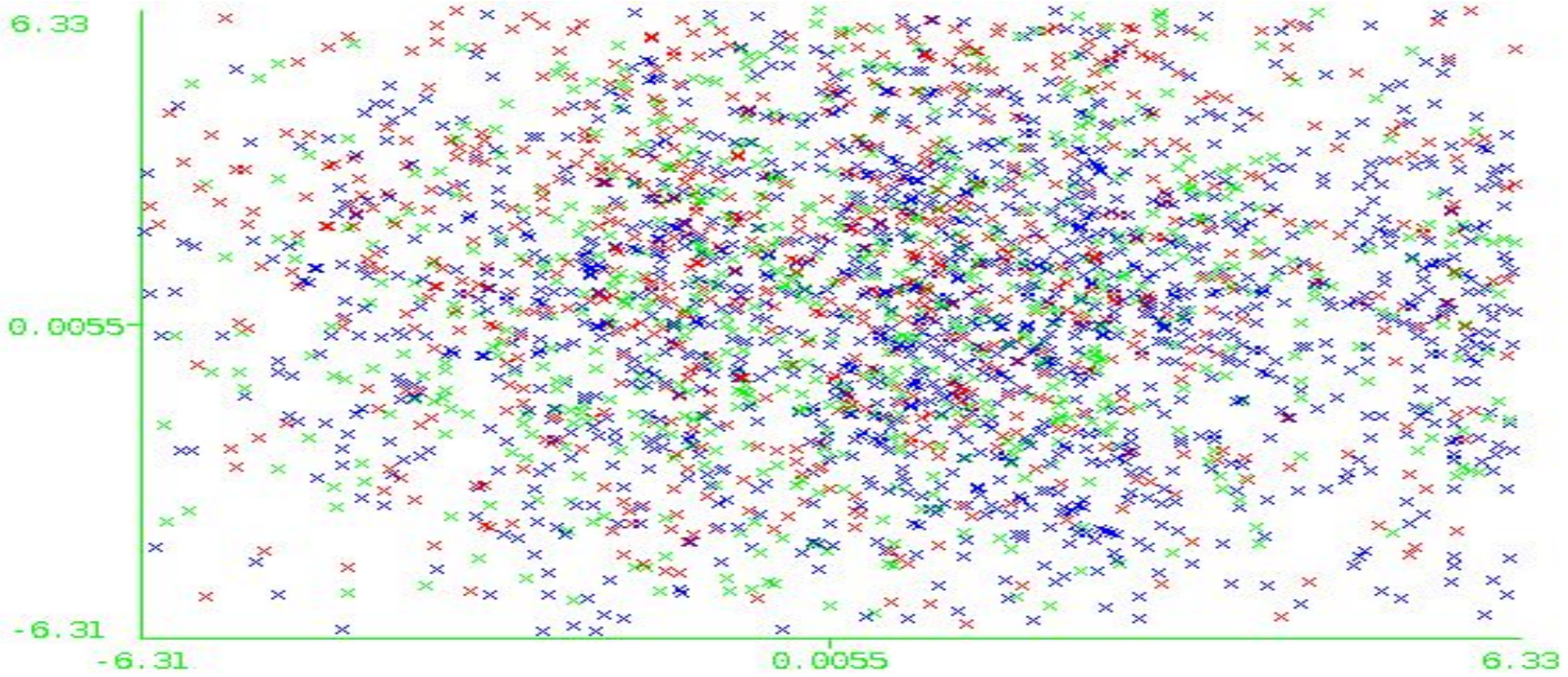
FEATURES: FORM

- If a team from group A wins against a team of group C (home of group C), points structure of Team A will be
- Points = ((+1) + (Ab * Ca)) * (1+Ad/100) (1)
- ((+1) + (0.15 * 0.4)) * 1.2
- And that of team C will be
- Points = ((-1) –(Ab * Ca)) * (1-Cc/100) (2)
- ((-1) – (0.15 * 0.4)) * 0.88

FEATURES : AQ,GOAL CONCEDED,CLEAN SHEETS

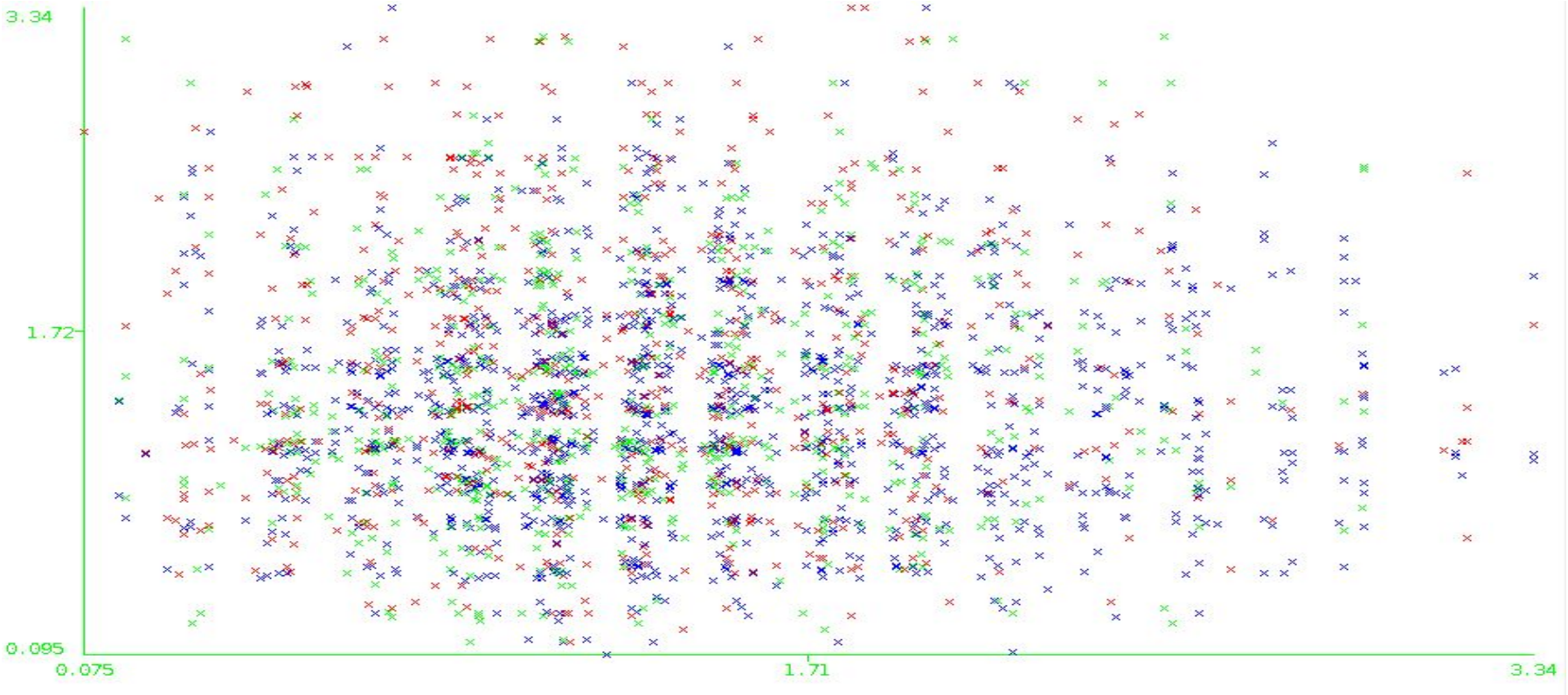
- AQ= $\sqrt{1/5 * \sum(\text{shots on target/total shots})^2} + \frac{1}{5} * \sum$ (goals scored)
- Clean sheets = ∑ clean sheets for 5 matches / 5
- Goals Conceded= ∑goals conceded/5

GRAPHS



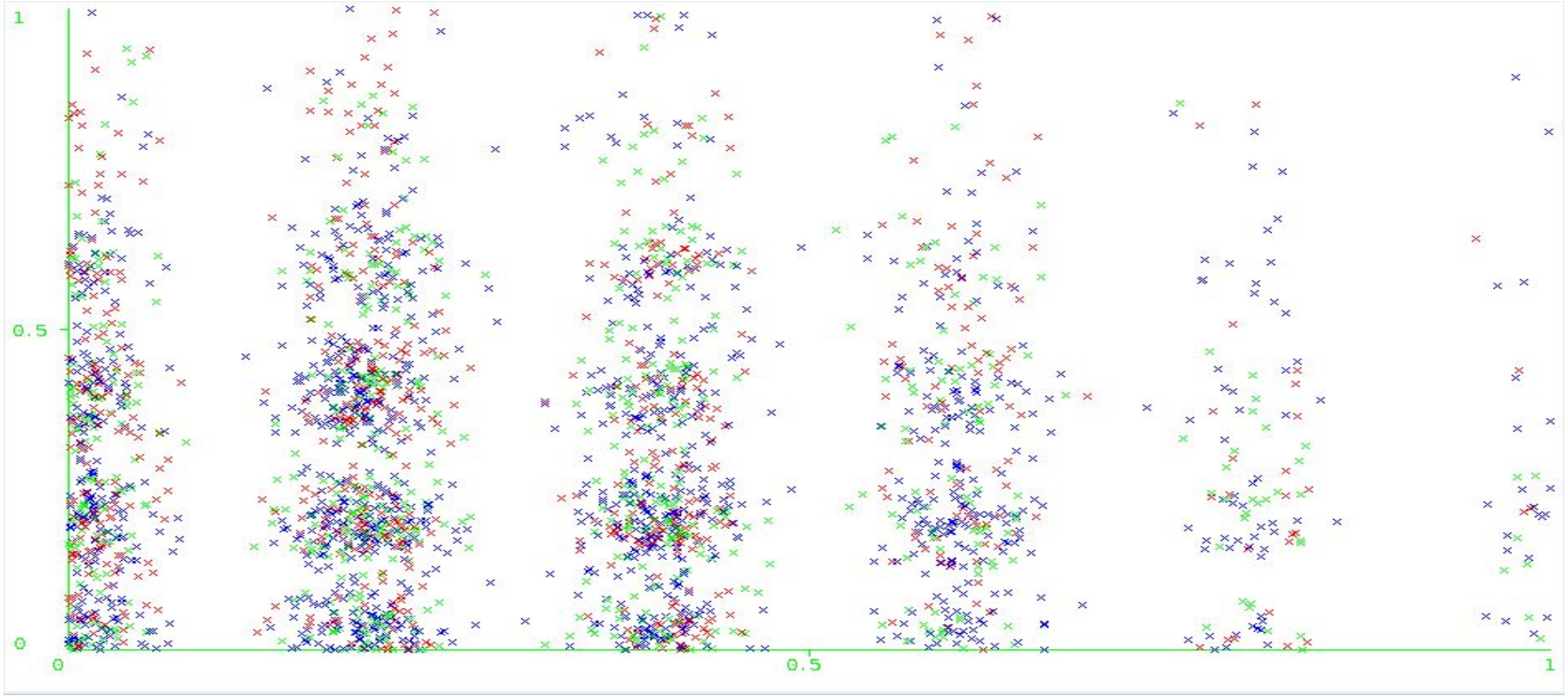
Form

GRAPHS



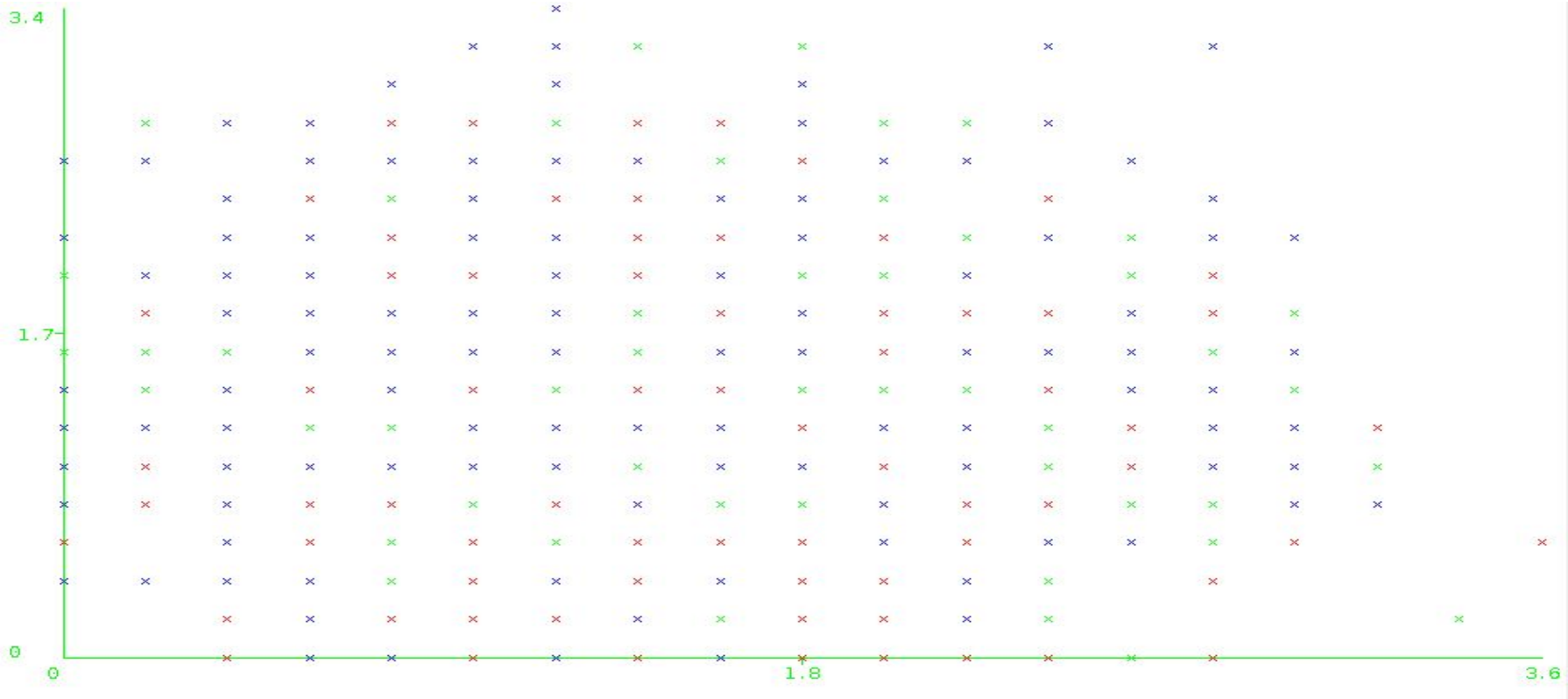
Attack Quotient

GRAPHS



Clean Sheets

GRAPHS



Goals Conceded

IMPLEMENTATION : CLASSIFIERS

- Implemented algorithms : Logistic regression and Vote algorithm (Naive Bayes and Random forest)
- Logistic Regression :

  Mean Accuracy : 0.6

  Characteristics :

  1. Supervised learning classification algorithm
  2. Only 2 classes classified (win and loss)

  Confusion matrix :

	Predicted win	Predicted loss	Predicted draw
Actual win	268	32	1
Actual loss	135	57	0
Actual draw	138	27	0

IMPLEMENTATION : CLASSIFIERS

- Vote algorithm (Naive Bayes and Random Forest) :

  Mean Accuracy : 0.63

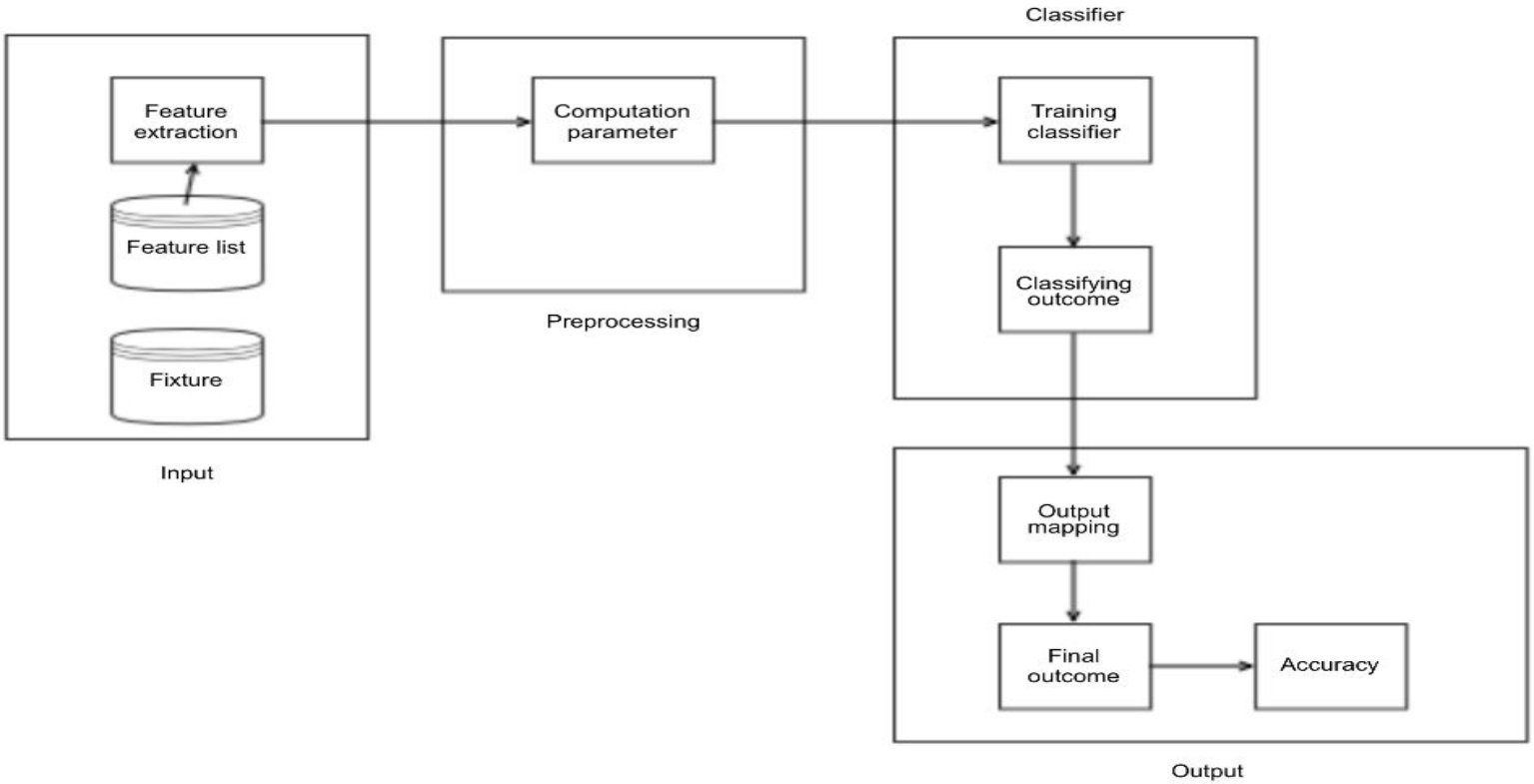
  Characteristics :

  1. Supervised learning classification algorithm
  2. All 3 classes classified
  3. Used ‘product of probabilities’ combination rule

  Confusion matrix :

	Predicted win	Predicted loss	Predicted draw
Actual win	235	52	14
Actual loss	114	66	12
Actual draw	112	44	9

SYSTEM ARCHITECTURE



System Architecture


Goal Accomplished !!!!