



# A Stochastic Decision Tree Model for Predicting The T20 International Cricket Match Result

Summit Haque, Moqsadur Rahman & Md. Saiful Islam  
Department of Computer Science and Engineering  
Shahjalal University of Science & Technology, Sylhet, Bangladesh.



# Overview

- Prediction in sports
- Cricket
- Data science in prediction
- Our approach
- Dataset & Preprocessing
- Factors
- Methodology
- Performance



# Prediction in sports

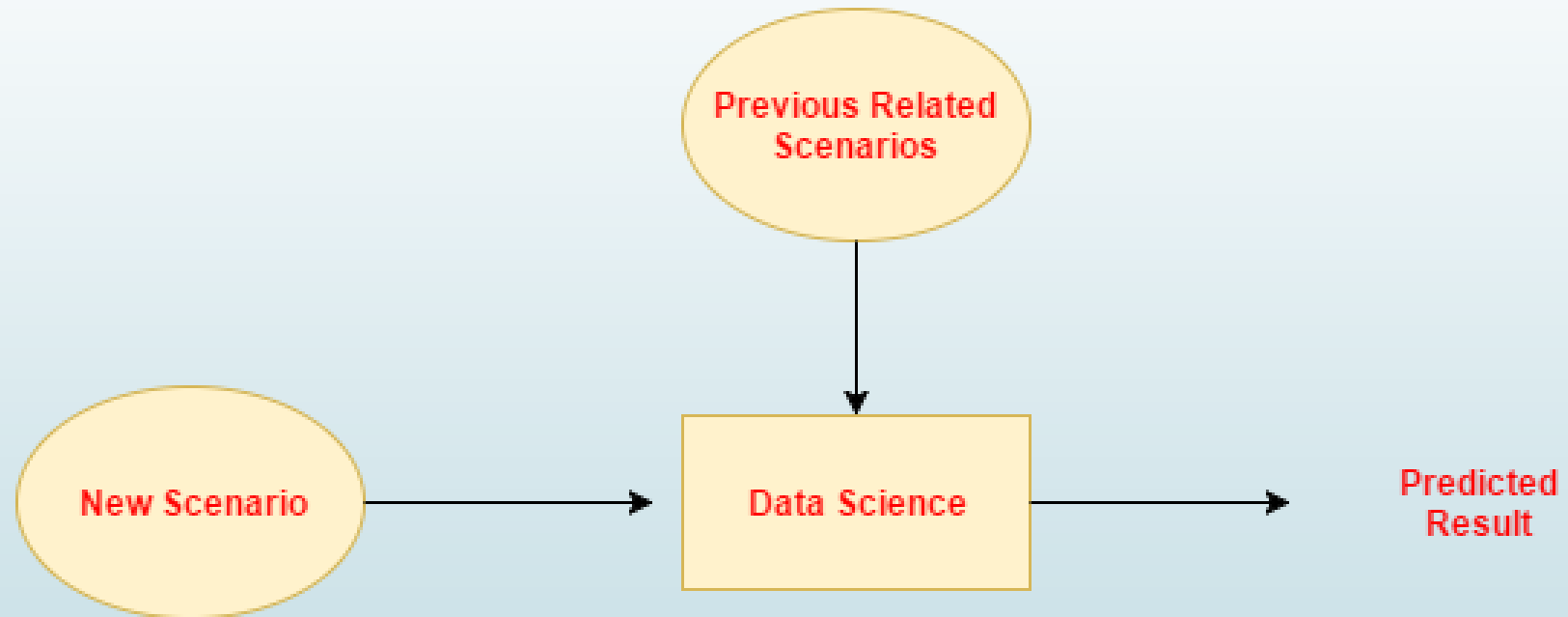
- Recent trend
  - Football, Cricket, Baseball etc.
- Types of predictions (example)
  - Before match
    - ✓ Result
    - ✓ Margin
  - Running match
    - ✓ Result
    - ✓ Margin
- Only result will be predicted



# Cricket

- A bat-and-bowl game
- Two teams with 11 players
- Two version
  - Longer: Test
  - Shorter: ODI, T20
- Two innings in shorter version.
  - First innings: Target is set
  - Second innings: Target is chased
  - Successful chase: Win
- T20 Format
  - Shorter version
  - Innings length 20 over(1 Over = 6 Balls)
  - More interesting

# Data science in prediction





# Our Approach

- Different approaches
  - Artificial Neural Network
  - Decision Tree
  - Naive Bayes
- Decision Tree
  - ID3
  - C4.5
  - CART
- We are using ID3 algorithm



# Dataset & Preprocessing

- Source: <http://cricsheet.org>
- T20 matches form 2005 to 2016
- 519 files in .yaml format
- Converted to .csv resulting in 504 files.
- Three sections
  - Training (304 files)
  - Validation (100 files)
  - Testing (100 files)

# .yaml Format

```
venue: The Rose Bowl
innings:
  - 1st innings:
    team: England
    deliveries:
      - 0.1:
        batsman: ME Trescothick
        bowler: B Lee
        non_striker: GO Jones
        runs:
          batsman: 0
          extras: 0
          total: 0
      - 0.2:
        batsman: ME Trescothick
        bowler: B Lee
        non_striker: GO Jones
        runs:
          batsman: 1
          extras: 0
          total: 1
      - 0.3:
        batsman: GO Jones
        bowler: B Lee
        non_striker: ME Trescothick
        runs:
          batsman: 0
```



# .csv Format

Ball	Batsman Run	Bowler	Date	Decision	Innin
0	0.1	0 B Lee	6/13/2005	bat	1st
1	0.2	1 B Lee	6/13/2005	bat	1st
2	0.3	0 B Lee	6/13/2005	bat	1st
3	0.4	0 B Lee	6/13/2005	bat	1st
4	0.5	0 B Lee	6/13/2005	bat	1st
5	0.6	0 B Lee	6/13/2005	bat	1st
6	0.7	2 B Lee	6/13/2005	bat	1st
7	1.1	0 GD McGrath	6/13/2005	bat	1st
8	1.2	0 GD McGrath	6/13/2005	bat	1st
9	1.3	0 GD McGrath	6/13/2005	bat	1st
0	1.4	0 GD McGrath	6/13/2005	bat	1st
1	1.5	0 GD McGrath	6/13/2005	bat	1st
2	1.6	0 GD McGrath	6/13/2005	bat	1st
3	1.7	1 GD McGrath	6/13/2005	bat	1st
4	2.1	4 B Lee	6/13/2005	bat	1st
5	2.2	1 B Lee	6/13/2005	bat	1st
6	2.3	0 B Lee	6/13/2005	bat	1st
7	2.4	4 B Lee	6/13/2005	bat	1st
8	2.5	4 B Lee	6/13/2005	bat	1st
9	2.6	1 B Lee	6/13/2005	bat	1st
0	3.1	0 GD McGrath	6/13/2005	bat	1st
1	3.2	4 GD McGrath	6/13/2005	bat	1st

# Preprocessed Data

B	C	D	E	F	G	H	I
Batting Avg All	Batting Avg Rec	Batting Ball By Wicket A	Batting Ba	Batti	Battin	Bowlin	Bowling
3.765567766	0.19047619	0.092307692	-1.48571	0.22	0.137	5.043	6.2157
-4.699488491	-3.351449275	-1.812276215	-0.99022	-0.2	-0.14	10.3	12.514
-4.262987013	-4.262987013	0.105194805	0.105195	-0.4	-0.37	11.27	11.274
3.650980392	1.181818182	0.742352941	0.054545	0.16	0.069	-1.057	2.6636
-0.851503759	-1.311111111	2.67406015	3.686667	-0.3	-0.39	-2.107	-4.132
-5.785947712	4.165441176	-3.589215686	1.956618	-0.1	0.074	11.3	7.4130
6.227513228	8.671122995	1.684126984	3.54385	0.23	0.219	-5.652	-9.0
1.845313157	10.41333333	1.950567465	8.176	-0	-0.02	4.711	-4.053
-0.161290323	-5.034210526	1.664516129	-0.21632	-0.1	-0.28	-4.373	-4.530
0.712566845	-4.34502924	0.110695187	-1.64912	0.04	-0.15	3.595	5.6256
5.287240729	8.986622074	1.817724701	6.6301	0.14	2E-04	3.291	2.1034
-2.282142857	-11.84897025	0.880714286	-3.5135	-0.2	-0.43	7.463	14.498
2.346077128	15.42592593	1.79900266	10.7619	0	0.023	7.174	-4.315
1.351830664	-2.531772575	0.152745995	-1.72074	0.08	-0.03	5.111	5.9666
6.331111111	13.48989899	3.093333333	6.663636	0.11	0.2	1.318	0.4148
-8.898644986	-15.27011494	-4.302113821	-8.21149	-0.2	-0.38	12.24	23.592
-0.631808279	5.176630435	0.423529412	6.11087	-0.1	-0.17	5.576	-5.333
-6.317375887	-10.66806723	-3.75035461	-7.63739	-0.1	-0.01	4.365	5.898
6.2	13.73099415	2.06775	10.29825	0.17	-0.05	0.737	-1.872
-0.410174881	5.6	-0.463593005	1	0.01	0.183	2.788	3.2408
-3.195488722	2.1	-0.393233083	2.29	-0.1	-0.05	-1.477	-6.744
2.029166667	0.706293706	2.08	2.307692	-0	-0.14	4.373	-6.784

# Factors

- Input to the model(feature)
- Prediction before the match
  - Difference of batting average  
(total runs taken / total wickets lost by team1 – total runs taken / total wickets lost by team2)
  - Difference of batting strike rate  
(total runs taken / total balls faced by team1 - total runs taken / total balls faced by team2)
  - Difference of bowling economy rate  
(total runs given / total overs delivered by team1 - total runs given / total overs delivered by team2)
  - Difference of bowling average  
(total runs given / total wickets taken by team1 – total runs given / total wickets taken by team2)

# Factors

- Difference of bowling strike rate  
(total balls delivered / total wickets taken by team1 - total balls delivered / total wickets taken by team2)
- Difference of wicket falling rate  
(total balls faced / total wickets lost by team1 - total balls faced / total wickets lost by team2)
- Face-to-face result  
(win percentage of team1 with team2 - win percentage of team2 with team1)
- Difference of recent win percentage  
(win percentage of team1 with all teams - win percentage of team2 with all teams)
- Home ground



# Factors

- Prediction of on-going match
  - Factors that are effective in prediction before match
  - Some new factors that are effective in running match
    - ✓ Difference of current and required run rate
    - ✓ Remaining wickets
    - ✓ Target remaining

# Methodology

## Decision tree

### ID3 Algorithm

- Information gain :  $\text{Gain}(S, F) = \text{Entropy}(S) - \sum \frac{|S_f|}{|S|} \text{Entropy}(S_f)$

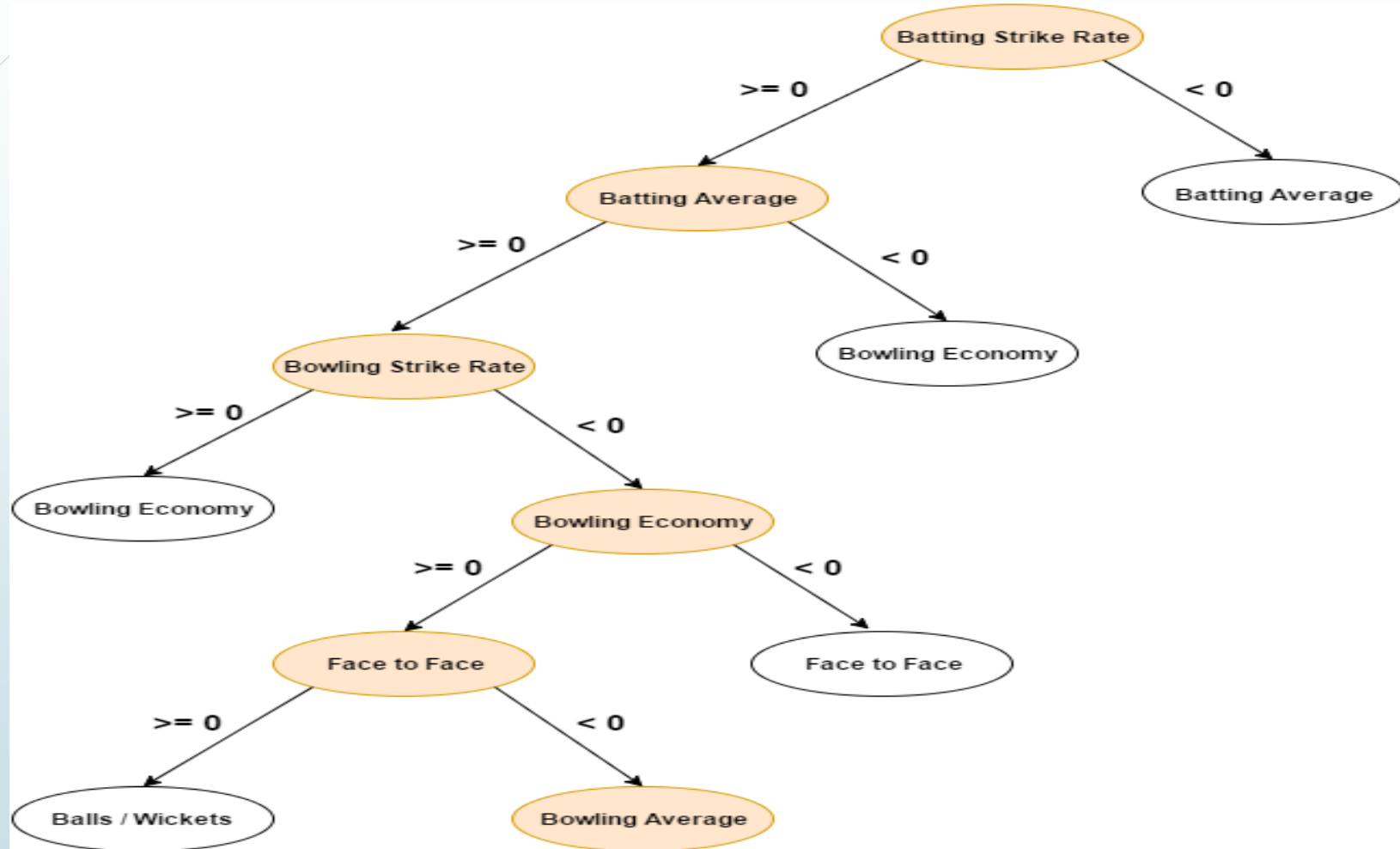
✓  $\text{Entropy}(p) = - \sum p_i \log_2 p_i$

- *F – difference of batting average, batting strike rate, bowling average, bowling economy, face to face etc.*

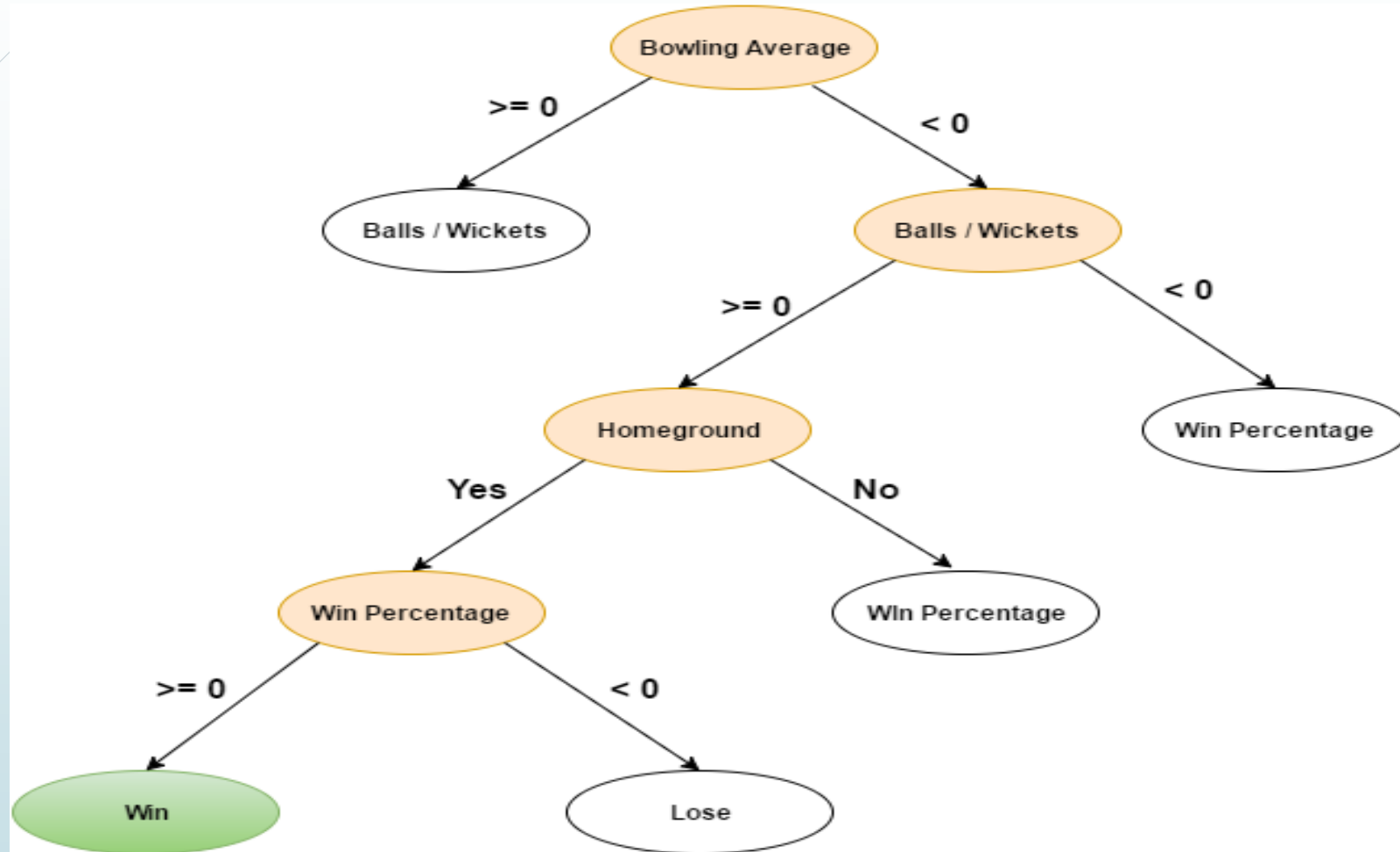
### For match before prediction, if difference of

- *Batting average: 3.28*
- *Batting strike rate: 5.67*
- *Balls / wicket: .89*
- *Bowling Average: -3.13*
- *Bowling strike rate: - 1.43*
- *Bowling economy: .58*
- *Face to face: -.2*
- *Win percentage: .34*
- *Home ground: yes*

# Methodology



# Methodology







# Methodology



## ► Ongoing

- Some extra features are added
  - Run rate – required run rate
  - wicket – required wicket
  - Target remaining
- The overall approach is same as the match before prediction

# Performance

- Total testing data: 100
- Before Match :  $\frac{56}{100} * 100 = 56\%$
- Ongoing Match :
  - First innings :
    - 1-5 over :  $\frac{53}{100} * 100 = 53\%$
    - 6-10 over :  $\frac{54}{100} * 100 = 54\%$
    - 11-15 over :  $\frac{59}{100} * 100 = 57\%$
    - 16-20 over :  $\frac{57}{97} * 100 = 58.77\%$

# Performance

## ➤ Ongoing Match:

### ▪ Second innings :

- 1-5 over :  $\frac{61}{100} * 100 = 61\%$
- 6-10 over :  $\frac{64}{100} * 100 = 64\%$
- 11-15 over :  $\frac{67}{96} * 100 = 69.79\%$
- 16-20 over :  $\frac{68}{89} * 100 = 76.40\%$

➤ Overall:  $\frac{539}{882} * 100 = 61.11\%$



THANK YOU