DISCOVERING STRONG PREDICTING ATTRIBUTES OF THE FIFA WORLD CUP RESULTS

ADAR GUY

JARED BARABASH

TATE NYCHYPORUK

TABLE OF CONTENTS

REPORT

INTRODUCTION

The FIFA World cup has grown from an attendance of 32,808 people per match, at its inauguration in Uruguay 1930, to 53,592 in Brazil 2014. With a greater stage, comes a greater variety of statistics, which make it harder to predict the outcome of the tournament. We've compiled statistics that were recorded from the beginning such as: goals against, goals for, penalty shot goals, wins, ties and losses since FIFA's origin. The more recent years were able to provide us with a much wider variety of data, including, but not exclusive to: shots missed, shots on target, corner kicks and completed passes. After an explanation of the methods used to extract the data, the necessary pre-processing needed to organize the data, and the data mining required to analyze the data, we will look at interesting trends and provide visual representation of mining tactics.

OBJECTIVES

The objectives of this data-mining project were as follows:

- Accumulate World Cup statistics and create a clean and useable data-set
- Analyze the set using data mining techniques and software
- Determine the statistics that best predict winning (and losing) teams
- Uncover trends in the statistics since the emergence or the World Cup
- Speculate on and report the acquired results

We hoped to develop an understanding of data scraping tools like R-Studio and use them in order to effectively and efficiently extract data that was important to us. Converting the data into functional datasets would give us insight into the importance of various variables associated with football. Gaining and understanding of the strength of each of these variables would be the final purpose of this study, and this would require understanding to depth the algorithms used by Weka. By scraping websites with R-studio and mining data with Weka, we aimed to create an interesting environment to process and analyze data.

DATA GATHERING

Data was found on-line in tables and charts like those seen below.

Figure 1.1

R	Team	G	P	W	D	L	GF	GA	GD	Pts.
1	Uruguay	3	4	4	0	0	15	3	+12	8
2	Argentina	1	5	4	0	1	18	9	+9	8
3	United States[63][64]	4	3	2	0	1	7	6	+1	4
4	Yugoslavia	2	3	2	0	1	7	7	0	4
Eliminated in the group stage										
5	Chile	1	3	2	0	1	5	3	+2	4
6	Brazil	2	2	1	0	1	5	2	+3	2
7	France	1	3	1	0	2	4	3	+1	2
8	Romania	3	2	1	0	1	3	5	−2	2
9	Paraguay	4	2	1	0	1	1	3	−2	2
10	Peru	3	2	0	0	2	1	4	−3	0
11	Belgium	4	2	0	0	2	0	4	−4	0
12	Bolivia	2	2	0	0	2	0	8	−8	0
13	Mexico	1	3	0	0	3	4	13	−9	0

Figure 1.2

NUMBER	TEAMS	GOALS SCORED	GOALS AGAINST	PENALTY GOAL	AVERAGE GOALS FOR	MATCHES PLAYED
1	Argentina	18	9	0	3.6	5
2	Uruguay	15	3	0	3.75	4
3	Yugoslavia	7	7	0	2.33	3
4	USA	7	6	0	2.33	3
5	Brazil	5	2	0	2.5	2
6	Chile	5	3	0	1.67	3
7	Mexico	4	13	1	1.33	3
8	France	4	3	0	1.33	3
9	Romania	3	5	0	1.5	2
10	Paraguay	1	3	0	0.5	2
11	Peru	1	4	0	0.5	2

EXTRACTION

We used R-studio to sift through tables provided by FIFA as well as Wikipedia. Built in functions like readHTML Table was used to simplify the process. This is the R code that was used to scrape and tabulate data from FIFA:

```
> XX <-READHTMLTABLE("HTTP://WWW.FIFA.COM/WORLDCUP/
ARCHIVE/CHILE1962/STATISTICS/TEAMS/GOAL-SCORED.HTML")
```

After running this code, R produced tables such as the one shown below. It can be seen that the conversion was not perfect and some additional pre-processing was required to clean it up.

Figure 1.3

```
NULL`
 numberâ-'â-¾     Teamsâ-'â-¾ Teamsâ-'â-¾ Goals scoredâ-'â-¾ Goals Againstâ-'â-¾ Penalty goalâ-'â-¾ Average Goals Forâ-'â-¾ Matches Playedâ-'â-¾
           1          Brazil         BRA                14                  5                 0                   2.33                    6
           2           Chile         CHI                10                  8                 1                   1.67                    6
           3      Yugoslavia         YUG                10                  7                 1                   1.67                    6
           4    Soviet Union         URS                 9                  7                 0                   2.25                    4
           5         Hungary         HUN                 8                  3                 0                      2                    4
           6  Czechoslovakia         TCH                 7                  7                 1                   1.17                    6
           7        Colombia         COL                 5                 11                 1                   1.67                    3
           8         England         ENG                 5                  6                 2                   1.25                    4
           9         Uruguay         URU                 4                  6                 0                   1.33                    3
          10      Germany FR         FRG                 4                  2                 1                      1                    4
          11          Mexico         MEX                 3                  4                 1                      1                    3
          12           Italy         ITA                 3                  2                 0                      1                    3
          13           Spain         ESP                 2                  3                 0                   0.67                    3
          14     Switzerland         SUI                 2                  8                 0                   0.67                    3
          15       Argentina         ARG                 2                  3                 0                   0.67                    3
          16        Bulgaria         BUL                 1                  7                 0                   0.33                    3
```

PRE-PROCESSING

Once the data was gathered, it was converted from space-separated values to CSV, then to the correct attribute-relation file format. To remove unwanted characters in the column titles, the built in Iconv function was used in R.

Figure 1.4

```
> REVISEDX <-DO.CALL(RBIND.DATA.FRAME, X)

> NAMES(REVISEDX) <- ICONV(NAMES(REVISEDX), TO='ASCII',
SUB='')
```

This left us with a cleaner table to work with as seen below.

Figure 2.1

number	Teams	Teams	Goals scored	Goals Against	Penalty goal	Average Goals For	Matches Played
1	Brazil	BRA	14	5	0	2.33	6
2	Chile	CHI	10	8	1	1.67	6
3	Yugoslavia	YUG	10	7	1	1.67	6
4	Soviet Union	URS	9	7	0	2.25	4
5	Hungary	HUN	8	3	0	2	4
6	Czechoslovakia	TCH	7	7	1	1.17	6
7	Colombia	COL	5	11	1	1.67	3
8	England	ENG	5	6	2	1.25	4
9	Uruguay	URU	4	6	0	1.33	3
10	Germany FR	FRG	4	2	1	1	4
11	Mexico	MEX	3	4	1	1	3
12	Italy	ITA	3	2	0	1	3
13	Spain	ESP	2	3	0	0.67	3
14	Switzerland	SUI	2	8	0	0.67	3
15	Argentina	ARG	2	3	0	0.67	3
16	Bulgaria	BUL	1	7	0	0.33	3

We then removed all the spacing from our Team names in order for an easy weka transition with use of the prebuilt data frame function in R Studio.

Figure 2.2

```
>AS.DATA.FRAME(APPLY(REVISEDX,2,FUNCTION(X)GSUB('\\S+',
'',X)))
```

Similarly we used this to remove all the suffixes on Country names such as Germany FR or Germany DR. When pulling data from Wikipedia, there were often multiple tables on a page. The data pulled required tables using keys, and were presented with tables similar to those seen below.

Figure 2.3

```
$`NULL`
                                R          Team    G    P    W    D    L   GF   GA   GD Pts.
1                               1        Brazil    3    6    5    1    0   14    5   +9   11
2                               2 Czechoslovakia   3    6    3    1    2    7    7    0    7
3                               3         Chile    2    6    4    0    2   10    8   +2    8
4                               4    Yugoslavia    1    6    3    0    3   10    7   +3    6
5  Eliminated in the quarter-finals         <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
6                               5       Hungary    4    4    2    1    1    8    3   +5    5
7                               6  Soviet Union    1    4    2    1    1    9    7   +2    5
8                               7  West Germany    2    4    2    1    1    4    2   +2    5
9                               8       England    4    4    1    1    2    5    6   -1    3
10    Eliminated in the group stage         <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
11                              9         Italy    2    3    1    1    1    3    2   +1    3
12                             10     Argentina    4    3    1    1    1    2    3   -1    3
13                             11        Mexico    3    3    1    0    2    3    4   -1    2
14                             12       Uruguay    1    3    1    0    2    4    6   -2    2
15                             13         Spain    3    3    1    0    2    2    3   -1    2
16                             14      Colombia    1    3    0    1    2    5   11   -6    1
17                             15      Bulgaria    4    3    0    1    2    1    7   -6    1
18                             16   Switzerland    2    3    0    0    3    2    8   -6    0
```

Unnecessary rows were removed and tables were processed in a similar manner.

Figure 2.4

```
>Z <- REVISEDX[-C(5),]
```

REMOVING UNWANTED ATTRIBUTES

At this stage we removed any unnecessary and unwanted attributes. The ISO tags were removed, as each team was already represented by its full name. Also, the number of matches played and the number of matches' won/lost/drawn were removed because they revealed too much about the placements of teams. This is because in the final round of the tournament a loss results in the team being knocked out. Therefore it is obvious that the team that played (or won) the most games won the tournament.

Similarly, the team that played the fewest games finished in last place. Because of their indicative nature, these statistics weren't at all interesting to analyze in the context of this project. Certain data that perhaps had not been sufficiently recorded at the time had to be accounted for in the dataset.

Finally, the numerical attributes (shots on goal, shots wide, etc.) had to be averaged over the number of games played in order to produce relative and comparable numbers for the different teams.

Once all the data was processed using these strategies, we were able to testing various algorithms to try and find interesting trends.

PERFORMANCE EXPERIMENTS

After the data set had been created, cleaned for erroneous data, and converted to a WEKA compatible .arff file, it was used to test the accuracy of each classifier. After some testing, the classifiers chosen were Naïve Bayes, J48, REPTree, and Perceptron. The reason behind this selection is because all four classifiers support binary, nominal, numeric and missing class attributes but the performance of each varies.

The cleaned data set contained 4 attributes (Year, Team, Corners, and Goals Against) with a class attribute (Placement). When conducting this performance experiment, the focus was on testing each classifier when changing the degree of the class attribute from binary (won, lost) to 14 places. With the results, four curves were graphed in order to gain visual representation of the decaying rate of precision. This would give insight into the strength of each classifier under the growing constraints. The results are represented in the table and graphs below.

Figure 3.1

Place	Naive Bayes	J48	Random Tr.	ML Percept.
2	94.46%	95.01%	92.24%	93.35%
3	89.75%	90.03%	84.21%	86.43%
4	83.66%	85.04%	77.29%	80.33%
5	79.22%	80.06%	70.36%	74.79%
6	74.24%	75.07%	65.65%	68.98%
7	71.47%	70.08%	58.73%	59%
8	66.48%	65.10%	54.02%	57.89%
9	61.77%	60.11%	47.65%	52.08%
10	56.79%	54.57%	43.77%	48.75%
11	51.52%	49.58%	39.61%	44.88%
12	46.26%	45.15%	35.18%	37.95%
13	40.44%	39.34%	29.64%	37.95%
14	36.29%	34.90%	24.38%	27.98%

Figure 3.2

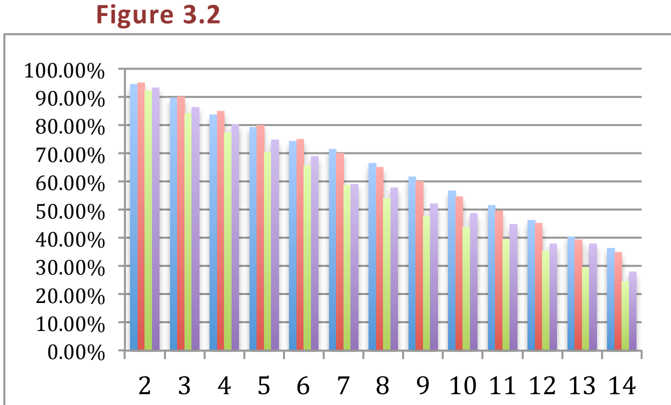
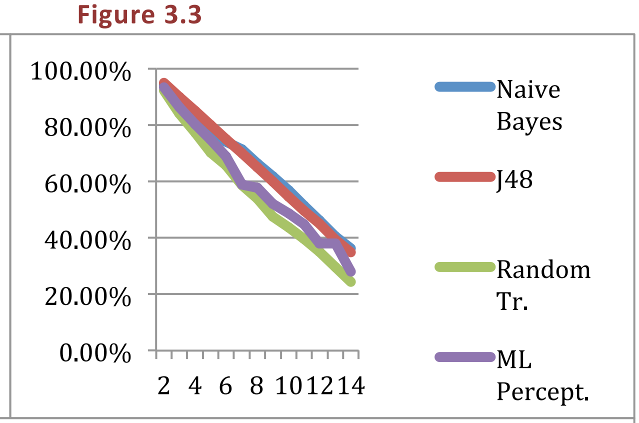
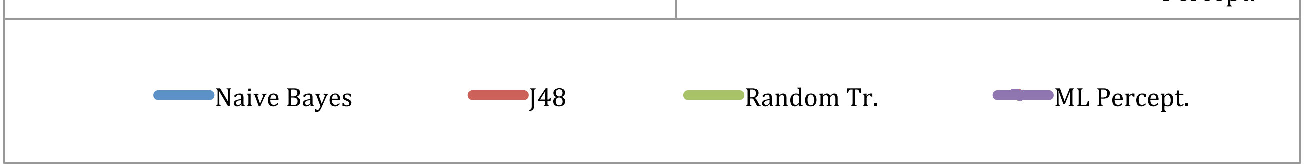


Figure 3.3





The process proved that J48, and Naïve Bayes were the strongest classifiers with very similar results between the two, while REPTree showed a more of an exponential decay and Perceptron showed some sporadic behavior. As the data had shown, the precision of each classifier declines drastically as more class values are introduced. In order to find a recall that maintained a high precision, the class values were to be organized into five brackets.

Figure 3.4

BracketA	BracketB	Bracket C	Bracket D	Bracket E
First	Fourth	Seventh	Tenth	Thirteenth
Second	Fifth	Eighth	Eleventh	None
Third	Sixth	Ninth	Twelfth	

TESTING EACH ATTRIBUTE

Due to the restraints on the data set, certain attributes had not been documented in earlier years, and this caused inconsistency in the data. The choice was made to break the dataset into a full dataset, and a subset of this data set.

The full dataset continued to use 4 attributes:

Figure 3.5

Year	Team	Corners	Goals Against

The full dataset was imported by Weka and mined using the same four classifiers as before (J48, Naïve Bayes, REPTree, and Perceptron). The results are represented in the table and graphs below.

Figure 3.6

Attribute	J48	Naïve Bayes	REPTree	Perceptron
Corners	43.29	48.71	45.18	41.88
GAAvg	43.29	49.41	46.35	43.29
ALL	42.82	48.94	46.12	40.47

Figure 3.7

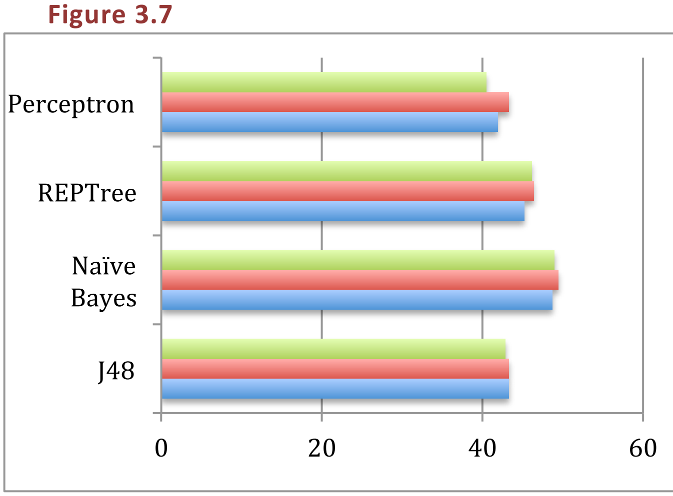
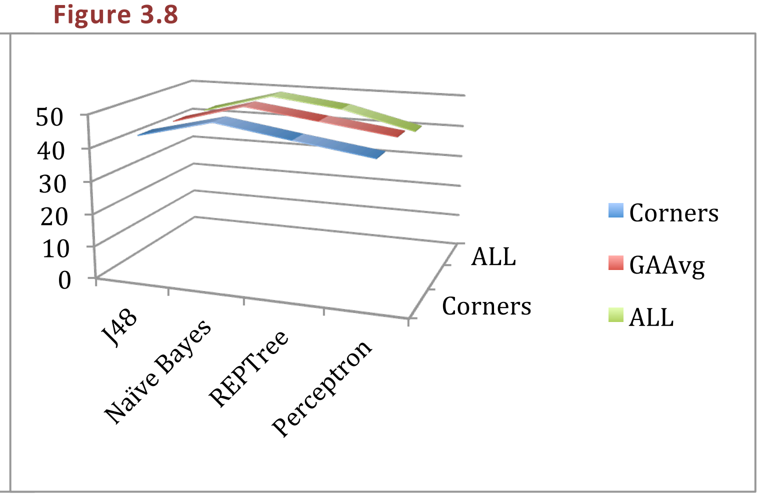


Figure 3.8



The subset dataset contained 10 attributes:

Figure 3.9

Year	Team	Goals For	Goals Against	Penalties	Shots On	Shots Wide	Free Kicks	Offside	Corners

The subset dataset, which contains data from 2002 onwards, contained more accurate statistics and was therefore able to include more attributes. This dataset was also imported through Weka and analyzed using the same classifiers to produce the results shown in the table and graphs below.

Figure 3.10

Attribute	J48	Naïve Bayes	REPTree	Perceptron
GFAvg	63.28	63.28	63.28	51.56
GAAvg	60.94	60.94	63.28	53.13
Penalties	63.28	63.28	63.28	44.53
ShotsOn	63.28	62.5	63.28	50
ShotsWide	63.28	61.72	63.28	47.66
FreeKicks	63.28	60.94	63.28	46.87
Offside	63.28	62.5	63.28	47.66
Corners	63.28	60.16	63.28	47.66
ALL	60.94	59.38	63.28	51.56

Figure 3.11

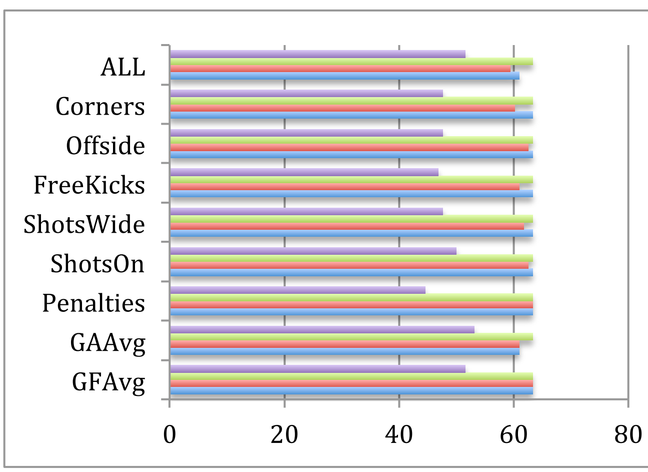
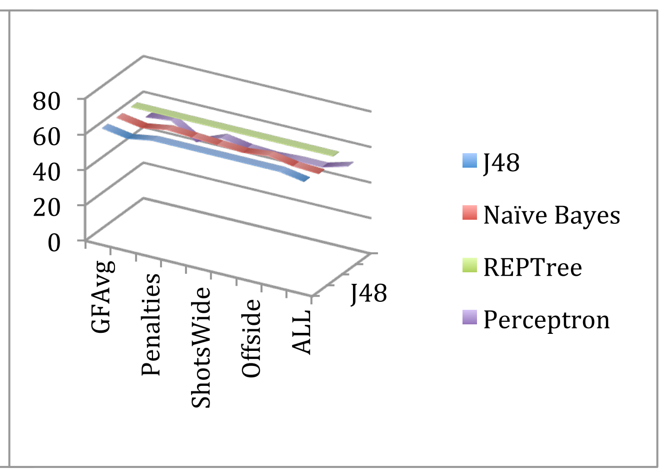


Figure 3.12



CONCLUSION

The results were very telling of which attribute most strongly predicts each outcome. Although the expectation was that “Goals For”, and “Goals Against” would be the strongest indicators, it was a surprise to find that in fact, “Penalties”, “Free Kicks” and “Corner Kicks” were equally as strong, if not stronger with certain classifiers. The quantitative results show that “Penalties” remain consistently high with each classifier (J48, Naïve Bayes, REPTree, Perceptron) and predicts as accurately as “Goals For” (63.28%). The lack of sufficient statistics for the World Cup is due to its short presence, but as more World cups take place, these trends should become more apparent.

APPENDIX

DESCRIPTION OF VARIABLES

Variable	Description
Team	Country (Team) name
Year	Year of the World Cup
G (Goals)	Total Goals Scored
P (Plays)	Number of Matches played
W (Win)	Number of Wins
D (Draw)	Number of Draws
L (Loss)	Number of Losses
GF (Goals For)	Goals Scored by Team
GA (Goals Against)	Goals scored against team
GD (Goals Compared)	Goals For – Goals Against
PTS (Points)	Number of Points (accumulated by wins and goals)
MP (Matches Played)	Amount of matches played
PEN (Penalties)	Number of Penalties incurred
SOG (Shots on Goal)	Amount of Shots on Goal by Team
CK (Corner Kicks)	Number of Corner Kicks
SW (Shots Wide)	Amount of Total Shots taken by Team
FK (Free Kicks)	Number of Free Kicks
OFF (Off sides)	Amount of Off sides

CLASSIFIER TESTING RESULTS

C = Correctly Classified, I = Incorrectly Classified

NAIVE BAYES

{YES,NO} = 94.46%(C)/5.54%(I)
{FIRST, SECOND, NONE} = 89.75%(C)/10.25%(I)
{FIRST, SECOND, THIRD, NONE} = 83.66%(C)/16.34%(I)
{FIRST, SECOND, THIRD, FOURTH, NONE} = 79.22%(C)/20.78%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, NONE} = 74.24%(C)/25.76%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, NONE} = 71.47%(C)/28.53%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, NONE} = 66.48%(C)/33.52%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NONE} = 61.77%(C)/38.23%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, NONE} = 56.79%(C)/43.21%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, NONE} = 51.52%(C)/48.48%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, NONE} = 46.26%(C)/53.74%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, NONE} = 40.44%(C)/59.56%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, THIRTEENTH, NONE} = 36.29%(C)/63.71%(I)

J48

{YES,NO} = 95.01%(C)/4.99%(I)
{FIRST, SECOND, NONE} = 90.03%(C)/9.97%(I)
{FIRST, SECOND, THIRD, NONE} = 85.04%(C)/14.96%(I)
{FIRST, SECOND, THIRD, FOURTH, NONE} = 80.06%(C)/19.94%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, NONE} = 75.07%(C)/24.93%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, NONE} = 70.08%(C)/29.92%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, NONE} = 65.10%(C)/34.90%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NONE} = 60.11%(C)/39.89%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, NONE} = 54.57%(C)/45.43%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, NONE} = 49.58%(C)/50.42%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, NONE} = 45.15%(C)/54.85%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, NONE} = 39.34%(C)/60.66%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, THIRTEENTH, NONE} = 34.90%(C)/65.10%(I)

RANDOM TREE

{YES,NO} = 92.24%(C)/7.76%(I)
{FIRST, SECOND, NONE} = 84.21%(C)/15.79%(I)
{FIRST, SECOND, THIRD, NONE} = 77.29%(C)/22.71%(I)
{FIRST, SECOND, THIRD, FOURTH, NONE} = 70.36%(C)/29.64%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, NONE} = 65.65%(C)/34.35%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, NONE} = 58.73%(C)/41.27%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, NONE} = 54.02%(C)/45.98%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NONE} = 47.65%(C)/52.35%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, NONE} = 43.77%(C)/56.23%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, NONE} = 39.61%(C)/60.39%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, NONE} = 35.18%(C)/64.82%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, NONE} = 29.64%(C)/70.36%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, THIRTEENTH, NONE} = 24.38%(C)/75.62%(I)

MULTILAYER PERCEPTRON

{YES,NO} = 93.35%(C)/6.65%(I)
{FIRST, SECOND, NONE} = 86.43%(C)/13.57%(I)
{FIRST, SECOND, THIRD, NONE} = 80.33%(C)/19.67%(I)
{FIRST, SECOND, THIRD, FOURTH, NONE} = 74.79%(C)/25.21%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, NONE} = 68.98%(C)/31.02%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, NONE} = 59%(C)/41%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, NONE} = 57.89%(C)/42.11%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NONE} = 52.08%(C)/47.92%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, NONE} = 48.75%(C)/51.25%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, NONE} = 44.88%(C)/55.12%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, NONE} = 37.95%(C)/62.05%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, NONE} = 37.95%(C)/62.05%(I)
{FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH, SEVENTH, EIGHTH, NINTH, TENTH, ELEVENTH, TWELFTH, THIRTEENTH, NONE} = 27.98%(C)/72.02%(I)

ATTRIBUTE TESTING RESULTS

FULL DATASET

J48 (42.82%(C)/57.18%(I))
CORNERS = 43.29%(C)/56.71%(I)
GAAVG = 43.29%(C)/56.71%(I)

NAIVE BAYES (48.94%(C)/51.06%(I))
CORNERS = 48.71%(C)/51.29%(I)

GAAVG = 49.41%(C)/50.59%(I)

REPTREE (46.12%(C)/53.88%(I))
CORNERS = 45.18%(C)/54.82%(I)
GAAVG = 46.35%(C)/53.65%(I)

MULTILAYER PERCEPTRON (40.47%(C)/59.53%(I))
CORNERS = 41.88%(C)/58.12%(I)
GAAVG = 43.29%(C)/56.71%(I)

SUBSET DATASET

J48 (60.94%(C)/39.06%(I))
GFAVG = 63.28%(C)/36.72%(I)
GAAVG = 60.94%(C)/39.06%(I)
PENALTIES = 63.28%(C)/36.72%(I)
SHOTSON = 63.28%(C)/36.72%(I)
SHOTSWIDE = 63.28%(C)/36.72%(I)
FREEKICKS = 63.28%(C)/36.72%(I)
OFFSIDE = 63.28%(C)/36.72%(I)
CORNERS = 63.28%(C)/36.72%(I)

NAIVE BAYES (59.38%(C)/40.62%(I))
GFAVG = 63.28%(C)/36.72%(I)
GAAVG = 60.94%(C)/39.06%(I)
PENALTIES = 63.28%(C)/36.72%(I)
SHOTSON = 62.5%(C)/37.5%(I)
SHOTSWIDE = 61.72%(C)/38.28%(I)
FREEKICKS = 60.94%(C)/39.06%(I)
OFFSIDE = 62.5%(C)/37.5%(I)
CORNERS = 60.16%(C)/39.84%(I)

REPTREE (63.28%(C)/36.72%(I))
GFAVG = 63.28%(C)/36.72%(I)
GAAVG = 63.28%(C)/36.72%(I)
PENALTIES = 63.28%(C)/36.72%(I)
SHOTSON = 63.28%(C)/36.72%(I)
SHOTSWIDE = 63.28%(C)/36.72%(I)
FREEKICKS = 63.28%(C)/36.72%(I)
OFFSIDE = 63.28%(C)/36.72%(I)
CORNERS = 63.28%(C)/36.72%(I)

MULTILAYER PERCEPTRON (51.56%(C)/48.44%(I))
GFAVG = 51.56%(C)/48.44%(I)
GAAVG = 53.13%(C)/46.87%(I)
PENALTIES = 44.53%(C)/55.47%(I)
SHOTSON = 50%(C)/50%(I)
SHOTSWIDE = 47.66%(C)/52.34%(I)
FREEKICKS = 46.87%(C)/53.13%(I)
OFFSIDE = 47.66%(C)/52.34%(I)
CORNERS = 47.66%(C)/52.34%(I)