Comprehensive Analysis of cases for Netball Australia

1. Table of Contents

Executive Summary

This report demonstrates exploratory and advance analysis of the dataset provided by the Netball Australia which includes information about matches, players, squads and their performance measures since 2009. Representing the best team of the world and being the most willing destination for the best player all over the globe for Netball, Netball Australia is interested in digging information for the past data which can be useful for increasing the competitiveness of Netball in Australia.

The report aims to explore the following issues on Australian national team:

- Efficiency of benchmark for choosing National Team Players
- Performance impact of split of Australia and New Zealand Championship
- Gaps (if any) in the new player substitution rules

For conducting exploratory analysis, the first phase of analysis tools such as Excel, Tableau and RapidMiner were used where the data were cleaned, analyzed and visualized to from hypothesis. Based on initially formed hypothesis, further advance analytics was performed where the hypothesis was tested, and models were formed using Excel and RapidMiner. The key findings of the combined analysis are presented below:

i. The average goal attempt of International matches is less than average goal attempt of domestic matches.
ii. The high number of goal attempt leads to the high number of goals of Australian Diamonds.
iii. The split of domestic league has made positive impact on the performance of Australian Diamonds.
iv. Most of the team started to use Substitution as key to hold game after 2014.
v. Most of the players are getting substituted between 300 to 600 seconds of each quarter to introduce fresh legs in the game.
vi. Squad "Fever" is mostly exploiting the rule of "substituting" the players for an injury.

Referencing to the findings of analysis, the report recommends not to base the selection of players of Australian Diamonds solely on SSN matches and to continue the split of domestic league which has provided Australian teams with a competitive edge the following to Netball Australia. Further, the report suggests using correct substitution of players to increase chance of winning and top level teams are to use rule of substitution correctly to ensure fitness players throughout the competition. Rolling substitutions for equal playing time is recommended with players swapped at any time. Similarly, to increase a chance of win best players must be substituted at right time and finally player interchanging rule could be introduced into SSN.

These findings could be useful to Netball Australia to assess their current situation, understand their strength and threats, while adopting these recommendations would contribute in eliminating the threat and exploring the opportunities.

1. Introduction and Approach

Started as Women's Basketball on Australia in 1897, the acceptance, growing demand, and unavoidable changes has given birth to existing unisex ball sport naming Netball Australia. Organizing the most advanced and professional domestic Competition in the World, Australia has succeeded to be most willing destination for best players over the globe. Before 2017, Australia and New Zealand used to share the domestic league. Australian National Team, Australian Diamonds, stands out to be the best of world and has victory around 82% of all its International Matches.

To lead in growing global competition and improve the level of performance, Netball Australia is curious about the following three issues and this report provides insights on:

1. What level of confidence in performance can Netball Australia have in the domestic competition to indicate the performance of those players at international tournaments?
2. How is the impact on the performance of Australian Diamonds after the split of Australia and New Zealand Championship?
3. What would be the new strategy of the coach to make changes in the court player?

This report mainly aims to analyze, and explore data provided by Netball Australia to comprehend its concerns undergoing two phases of analysis: Exploratory and Advance analysis. In Exploratory analysis, statistics provided about matches, players, squads and their performance measures since 2009 is analyzed, cleansed and visualized using Excel, Tableau and RapidMiner. Based on the data visualization hypothesis has been developed which has worked as steppingstone for further analysis. Similarly, in Advance analysis the hypothesis has been tested and prediction model has been generated using Excel and RapidMiner respectively. Thus, the report tends to discover the pattern underlying in the given data and draw a logical recommendation to Netball Australia to analyze and improve the performance of players and squad.

2. Assumptions

The Exploratory and Advance Analytics, are subject to the following assumptions:

1. The year mentioned in the column 'Match_id" stands for the year the tournament was held.
2. Squad represented with name "Australian", "Australia", "Diamonds" and "Australia Diamonds" stands for the national team of Australia which is "Australian Diamonds" with squad ID "811"
3. Goals, penalties, gains, feeds and goal attempts are Key Performance Indicator of players and teams
4. Bad Passes, Bad Hands, Offsides, breaks, turnovers and Missed Goal are Key Indicator of Performance Error.
5. Missing values and errors in data were removed to increase accuracy of evaluation.
6. There is multi-collinearity among attributes selected in the models (expect with label).
7. A new attribute has been generated to fit models.
8. For Hypothesis 1-3 for model building 70% of player and team data was used as input to train the model and remaining 30% was used for evaluating the model performance.

3. Data Analysis

In this section exploration on the three concerns of Netball Australia is presented by developing 2hypothesis for each concern. Analysis is conduced using explorative approach followed by advanced analytics and finally interpreting the findings.

3.1 Evaluation of using SSN as a yardstick

Netball Australia uses the performance of players in SSN as a criterion to select players for Australian Diamonds. Hypothesis have been developed to compare the performance of players in domestic matches and international matches.

3.1.1 Hypothesis 1: In comparison to SSN, the confidence of players to attempt goal in IR decreases by 50%.

Fig 1: Performance comparison of players on SSN and IR

Initially exploration presented average feeds, gains, goal attempts, intercepts and penalties to have decreased in international round than in SSN for Australian players showing significant reducing of 50% goal attempts in international rounds compared to SSN.

Furthermore, data of 28 players playing for Australian Diamonds were considered for both SSN and IR. The data was split into SSN and IR and average goal attempts of SSN Matches was evaluated to be 9.52. Hypothesis testing was performed using z test as shown in appendix for the deduced hypothesis creating null and alternate hypothesis.

- Null Hypothesis: The average goal attempts of International matches is equal to average goal attempts of SSN matches
  Ho: $\mu = 9.52$

- Alternate Hypothesis: The average goal attempts of International matches is less than average goal attempts of SSN matches
  H1: $\mu < 9.52$

p value obtained for the test was less than 0.05 leading to reject null hypothesis proving average goal attempts of the players in international match is less than average goal attempts in domestic tournaments.

Additionally, weight by correlation showed that goal attempts is highly correlated to goals scored so predictive model using linear regression was created with goals as labels and goal attempts as predictors to know the performance of players in international matches based on SSN.

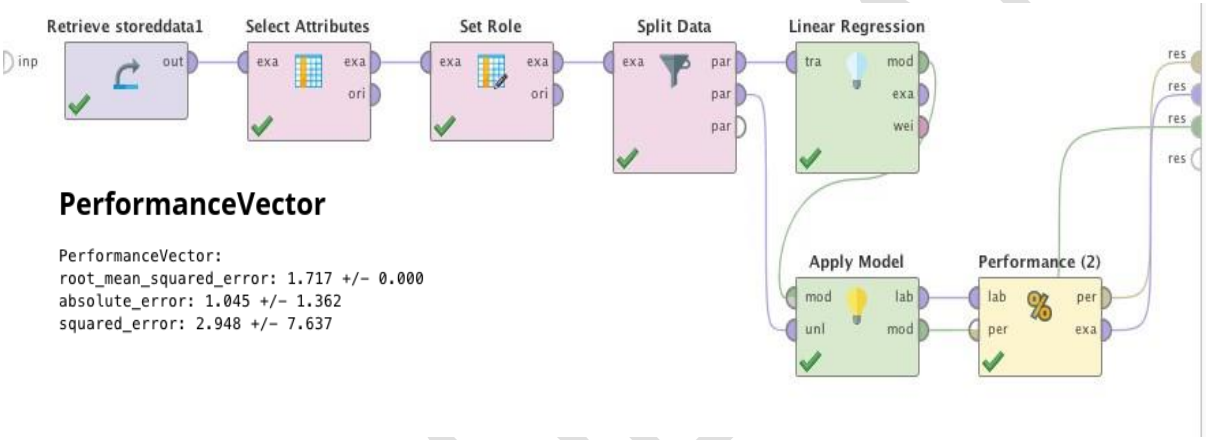


Fig 2: Predictive model for hypothesis 1 along with performance vector

The model can further be used by Netball Australia to evaluate the performance of players in International Matches based on the domestic match. However, it only predicts the goal within a range of +/- 1.7 goals.

It was observed that due to performance being decreased in international matches than SSN using it as a yardstick might raise some concerns.

3.1.2 Hypothesis 2: In comparison to ANZC, the confidence of players to attempt goal in IR decreases by 35%

Fig 4: Performance comparison of players on ANZC and IR

The dashboard above shows the performance of 28 Australian Diamond players who have played at ANZC and internationally as well. Australian players performance in key performance indicators such as average penalties, average intercepts, feeds and goal attempts have underperformed at international level than ANZC.

Furthermore, null hypothesis and alternate hypothesis were created to test the above hypothesis

- NULL HYPOTHESIS: Average goal attempts by Australian Players in ANZC tournament is equal to the average goal attempts by the Australian players in International matches.
  Ho: $\mu = 8.98$
- ALTERNATE HYPOTHESIS: Average goal attempts by Australian Players in International matches is less than average goal attempts by the Australian players in ANZC tournament.
  H1: $\mu < 8.98$

We reject Null hypothesis as p value obtained for the test was less than 0.05. Thus, there is a decrease of 36.8% in average goal attempts at international matches in comparison to average goal attempts at ANZC tournament.

Moreover, predictive model using linear regression classifier were created. Dataset consisting of ANZC data & international matches was used separately. Our model predicts the goal attempts by players at International matches and at ANZC tournaments. It was noticed that the goal attempts have the highest correlation.

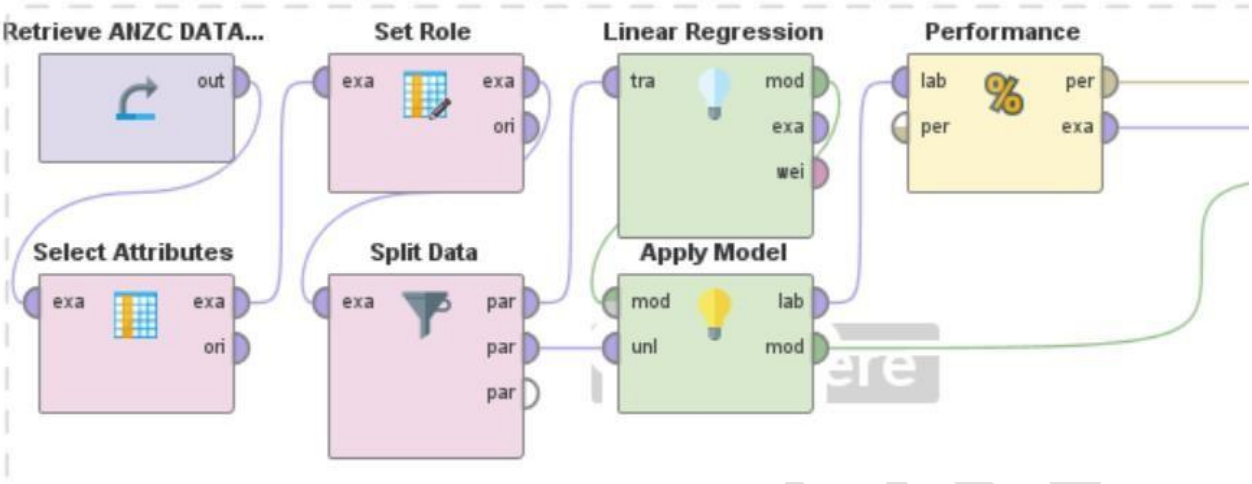


Fig 4: Predictive model for hypothesis 2

Model created using ANZC data predicts the goals within range of +/- 1.5goals and root mean squared error, absolute error, relative error, squared correlation and correlation are taken as performance measures. Similarly, another model was created using data from international matches whose performance is shown in appendix.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.525 +/- 0.000
absolute_error: 0.829 +/- 1.279
relative_error: 8.71% +/- 7.78%
correlation: 0.995
squared_correlation: 0.990
```

Row No.	goalAttempts	predictio... ↓	goals	playerId
469	57	58.515	50	80113
535	54	58.515	50	80113
138	53	57.348	49	80113
198	55	57.348	49	80113
84	52	52.679	45	80113
458	47	51.512	44	80113
293	47	50.345	43	80052
58	50	49.178	42	80113
221	45	46.843	40	80072
454	44	46.843	40	80052
152	41	45.676	39	80113
273	43	45.676	39	80113

Fig 5: Performance Vector and Example Set of predictive model for performance of players on ANZC

3.2 Performance Evaluation of Australian Diamond after the split of ANZC into SSN and ANZP

Netball Australia got separated from New Zealand and launched its own domestic competition in 2017. This hypothesis examines if the split has really benefitted Australian Diamonds in terms of their performance. This hypothesis also analyses impact on the performance of Australian Diamonds after domestic competition split.

3.2.1 **Hypothesis 3**: Average Gain has been increased by 5times after the split of domestic championship of Australia and New Zealand

Fig 6: Performance Evaluation of Australian Diamonds after split

From the exploratory analysis performed it was observed that post-split, some performance metrics like Goals, penalties, gains, feeds and goal attempts seemed to follow a positive trend.

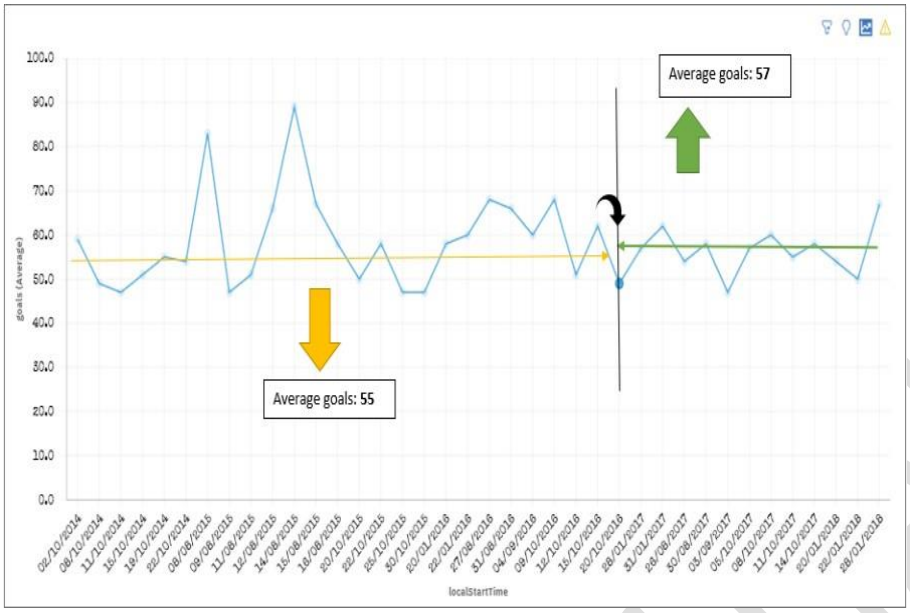


Fig 7: Average Goal Evaluation before and after split

Following observation were observed:

- Average number of goals by Australian Diamonds used to be 55 before split.
- Average number of goals scored increased by 3.63% to 57 goals after split.
- Out of all the metrics, Average gains witnessed the highest spike of about 13%.
- Average penalties increased by 6% from 61 to 65.
- Average feeds increased by 5.18% after the split from 74 to 78.
- Average goal attempts post-split observed 1.76% increase from 63 to 64.

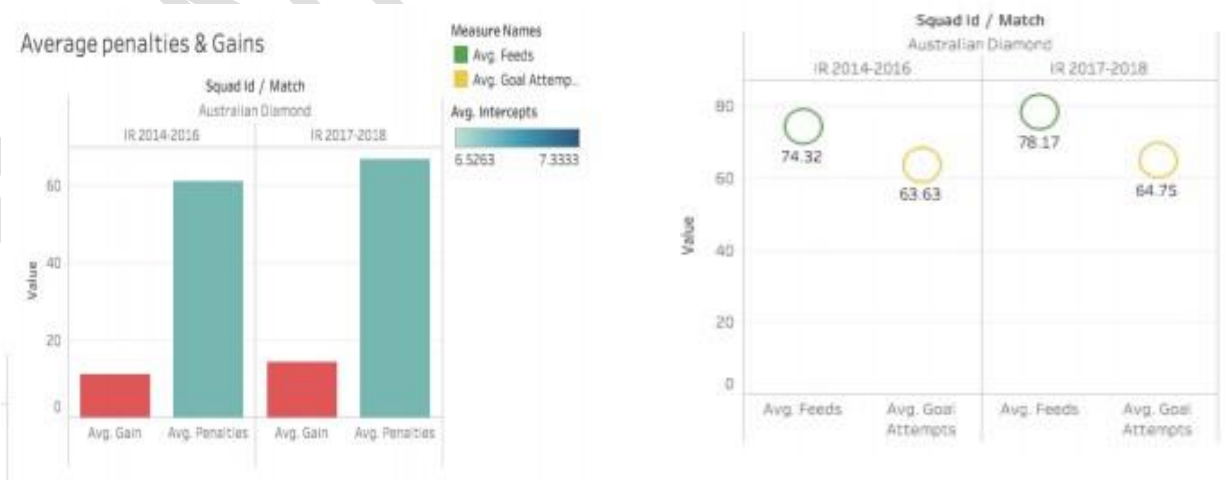


Fig 8: Average penalties, gains, feeds and goal attempts before and after split

As part of further research interest, a predictive model using neural networks was created in rapid miner to predict the number of goals scored by Australian diamonds. The Hyper parameters used for creating the neural networks were 200 training cycles, 0.01 learning rate and 0.9 momentum.

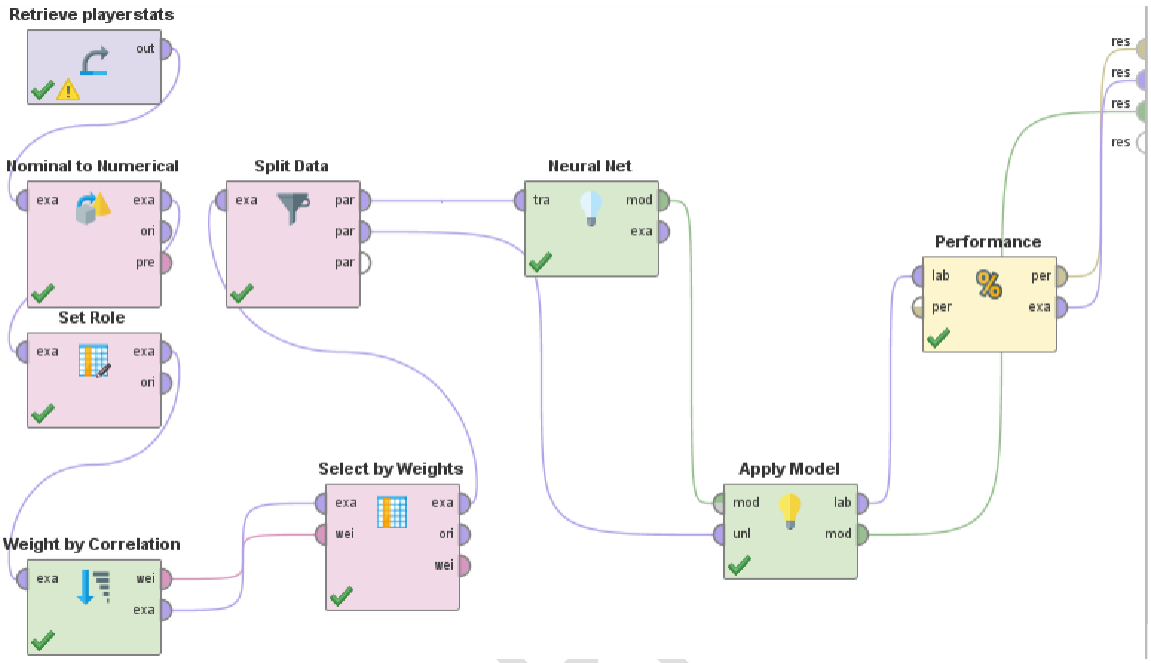


Fig 9: Predictive model for hypothesis 3

From the correlation matrix generated it was clear that number of goals scored was highly correlated to number of goal attempts followed by number of goal misses. On the other hand, starting position code, current position code and offensive rebounds had a decent influence on the goals scored.

goals	prediction(goals)	startingP...	currentPosit...	offensiveRe...	goalMisses	goalAttempts
3	3.014	0	1	1	2	5
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
24	24.429	0	0	2	4	28
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0
0	0.013	0	0	0	0	0

attribute	wei... ↓
goalAttempts	0.988
goalMisses	0.602
startingPositionCod...	0.578
currentPositionCod...	0.506
offensiveRebounds	0.499

Fig 10: Weight by correlation matrix

The model created has a correlation of 0.9 which not only shows that it's a good fit for the given data but also signifies dependency of the independent variables like goal attempts, goal misses etc., on goals scored. The model is pretty good with predictions with an accuracy of over 85% making negligible absolute and root mean square errors.

absolute_error

absolute_error: 0.081 +/- 0.229

root_mean_squared_error

root_mean_squared_error: 0.243 +/- 0.000

Fig 11: Error for predictive model after the split

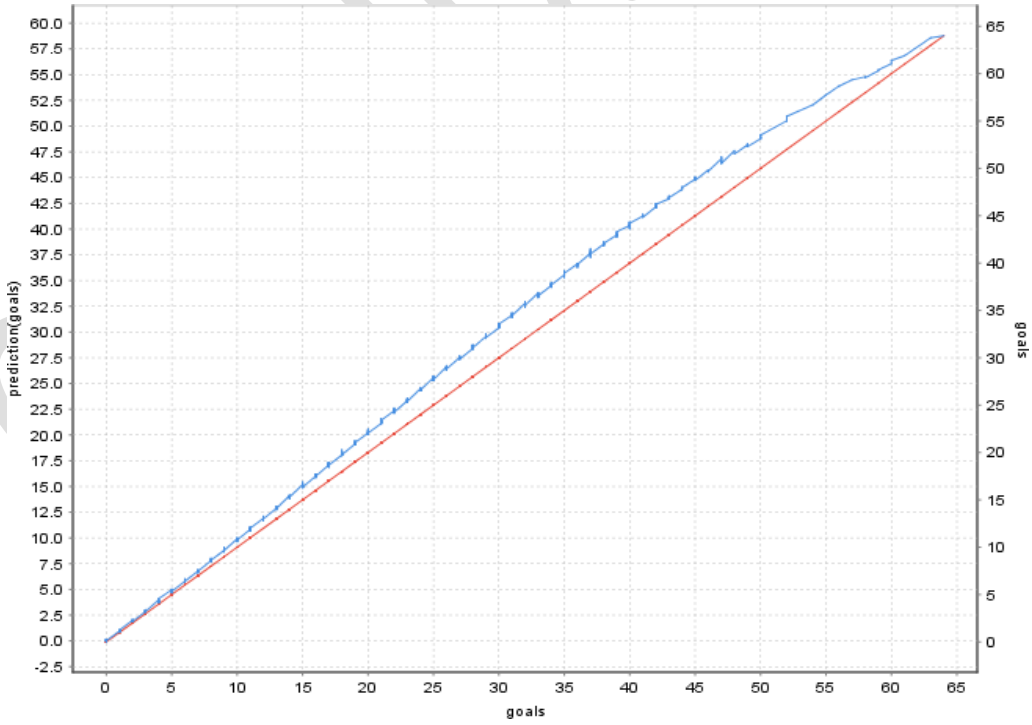
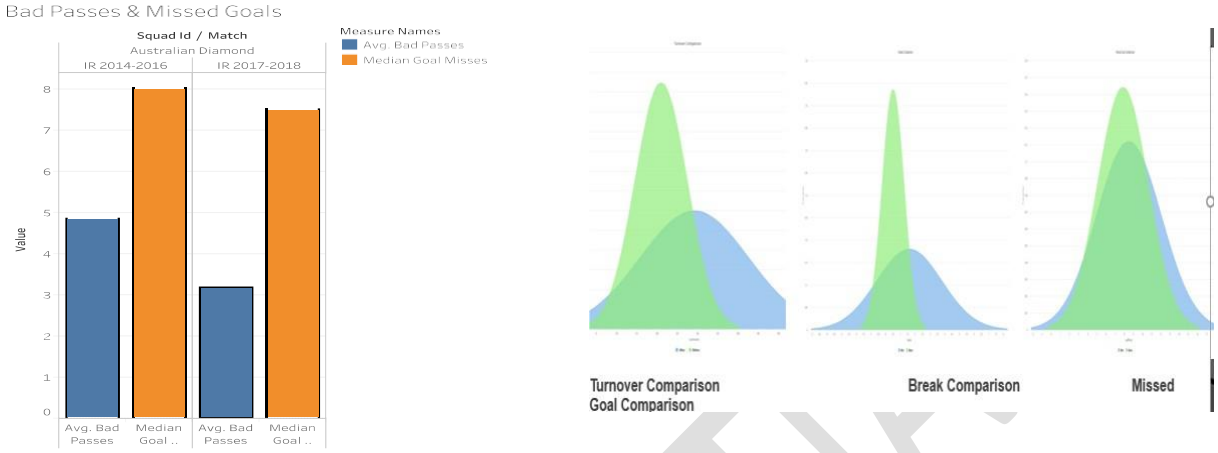


Fig 12: Actual vs Predicted value from the model

3.2.2 **Hypothesis 4**: The Number of Errors of Australian Diamonds has decreased by 20% after the split of domestic championship of Australia and New Zealand.



A *B*

Fig 13: A: Changes in number of errors in performance of Australian Diamonds after split

B: Bell curve distribution of parameters before and after split

For analyzing changes in performanceerrors of Australian Diamonds, six statistics missed goals, turnover, breaks, offsides and bad passes were considered as key indicator and records of 38 international matches was derived. The comparative exploratory analysis of data before and after 2017, showed declination of those factors by 50%.

With the necessity of advance analytics for the better picture, a prediction model was formed using K-nn Model. Available dataset was spilt in 8:2 ratio for developing and testing model respectively. Also, a new attribute "Match Timing" was created to segregate data before and after 2017. Predictive and accuracy power of model varies on the label, attributes, and parameters chosen. Considering least error of the model, break was considered as label and prediction model was performed.

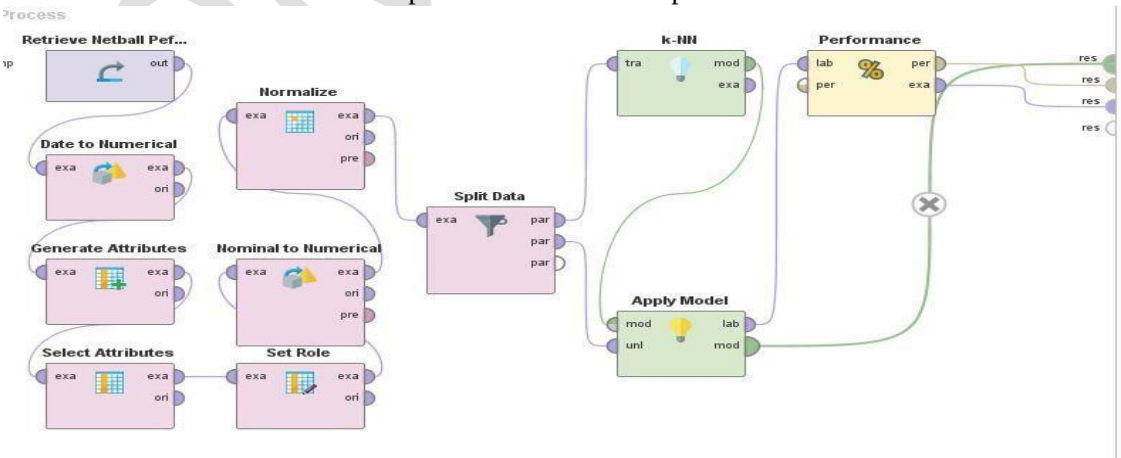


Fig 14: Predictive model for hypothesis 4

breaks	prediction(b...	Match Timin...	Match Ti...	offsides	badPasses	badHands	goalMisses	turnovers
1	0.927	1.651	-1.651	-0.526	-0.084	-0.295	-0.346	4.493
0	0.894	1.651	-1.651	0.029	0.979	0.328	-0.009	-0.661
0	0.969	1.651	-1.651	-0.526	-0.615	-0.295	-1.696	0.305
1	0.916	-0.590	0.590	-1.082	-0.084	0.328	-0.009	-0.446
0	0.773	-0.590	0.590	1.696	1.510	0.950	1.341	-0.446
0	0.918	-0.590	0.590	1.140	0.447	0.328	-1.021	-0.232
1	0.666	-0.590	0.590	-1.082	-0.615	-1.540	1.003	-0.661
1	0.917	-0.590	0.590	-1.082	-0.084	0.328	-0.346	-0.876

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.501 +/- 0.000
absolute_error: 0.501 +/- 0.008
normalized_absolute_error: 1.002
root_relative_squared_error: 1.002
squared_error: 0.251 +/- 0.008
prediction_average: 0.500 +/- 0.500
```

Fig 15: Performance Vector for the predicted model

Root mean square error of model is 0.501 which shows standard deviation of prediction errors or residuals from the regression line. Thus, prediction errors are moderately spread out. The model predicts value of break to be 0.927 when actual value is 1. But when actual value is 0 it predicts it to be 0.894. Thus, accuracy of the model is only around 50%. However, due to lack of extra set of unseen data, model deployment can be biased and unproductive.

The combination of exploratory and advance analytics demonstrates that the performance errors of Australian Diamonds has decreased in some extent while the possibility predicting of such errors is only around 50%. Thus, current performance of Australian Diamonds advocates the pros of the split of domestic league and relaxation in the restriction of having team members from clubs competing nation in SSN.

3.3 Patterns indicating change in court players by coach

3.3.1 Hypothesis 5: Most of the coaches are mainly substituting their players in mid of the period in the range of 350-600

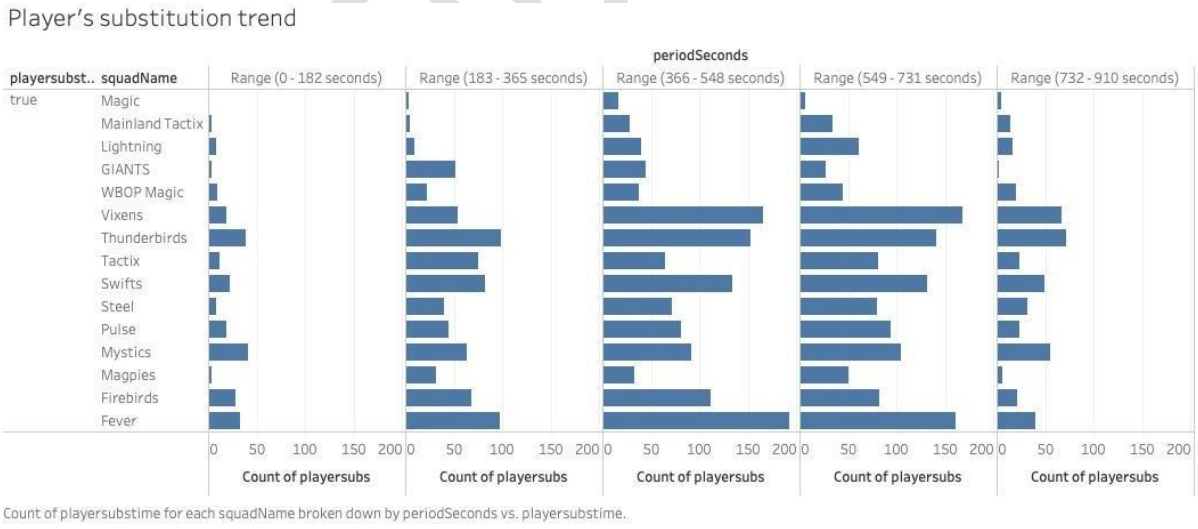


Fig 16: Distribution of players substitution according to periods

The exploratory analysis tried to explain how netball team coaches are using substitution rule for interchanging players during the game and in what duration they are taking advantage of this rule. Five different bins were created for examining about the timing of substitution of players for analyzing about the trend used by coaches for substitution.

Based on exploratory analysis, model in rapid miner have been created to understand the timing scenario of player substitution during game. Below are dataset and parameter details used for our model.

Dataset – Teams, Player_subs and Teams_stats

Label - Periodseconds

Predictors – Frompos and topos

team and Player_subs data have been joined for observing hypothesis. But we have used team stats and player subs process for generating new variable called as "playersubstime" for which we created five different bins of time when player was substituted for observing better results.

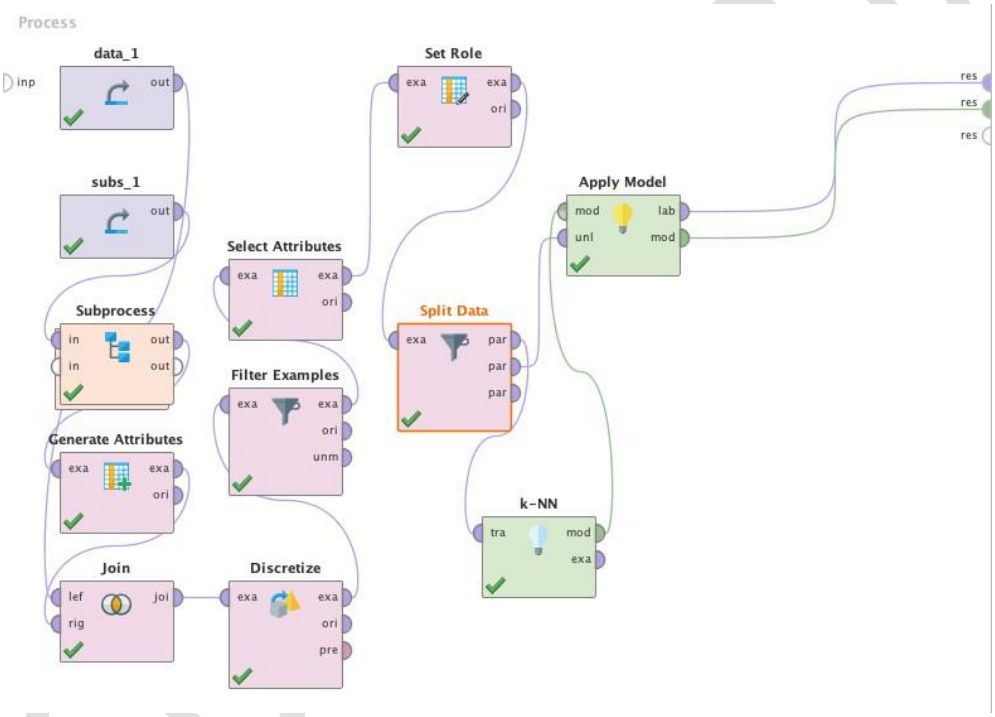


Fig 17: Predictive model for hypothesis 5

Matches have been grouped based on "year" when they were played as shown in subprocess of our model. We performed optimization on data and best accuracy and results were observed using K-nearest neighbors using Euclidean measure and having K value as 10.

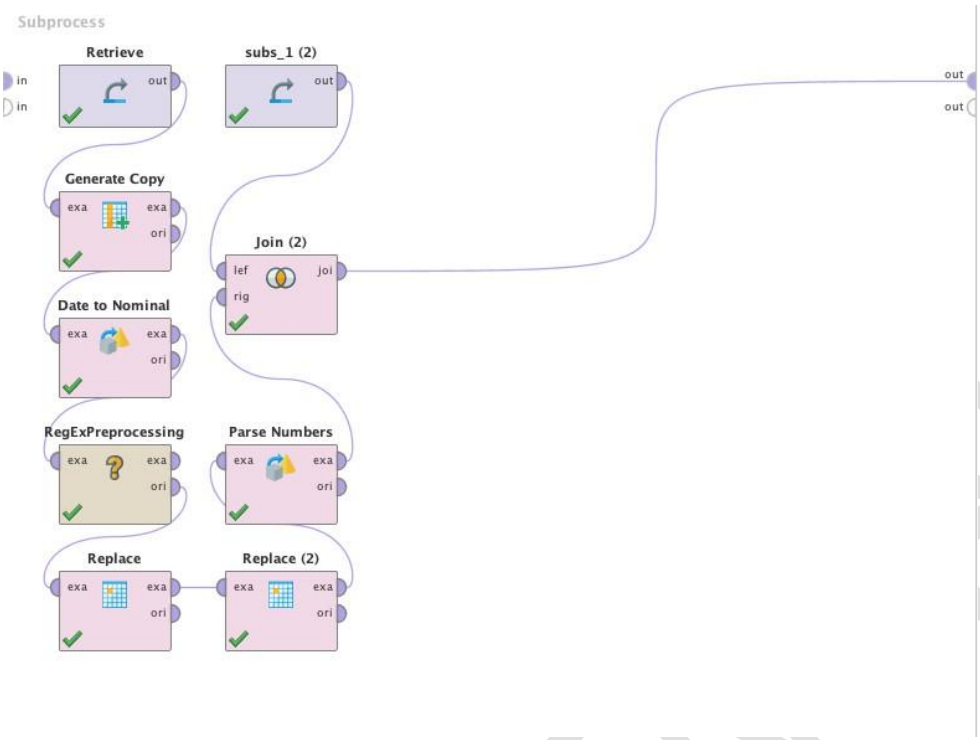


Fig 18: Sub process for grouping matches by Years

- It was observed that most of the players are getting substituted between 300 to 600 seconds of each quarter to introduce fresh legs in the game.
- Also, after 2nd quarter i.e. when half of the game is completed teams are using most of the substitution of players as well.
- Squad "Fever" is mostly exploiting rule of "interchanging" players for an injury followed by squad "Vixens" and so on.

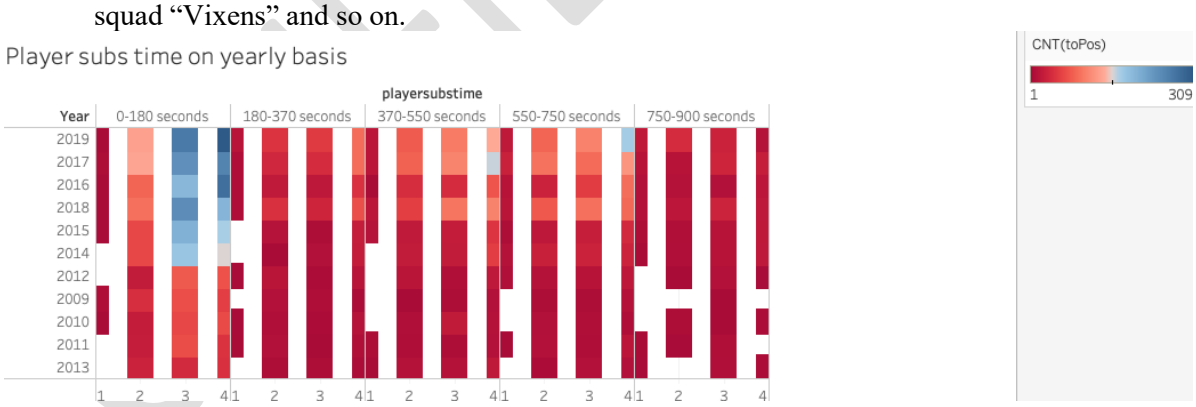


Fig 19: Player substitution time on yearly basis

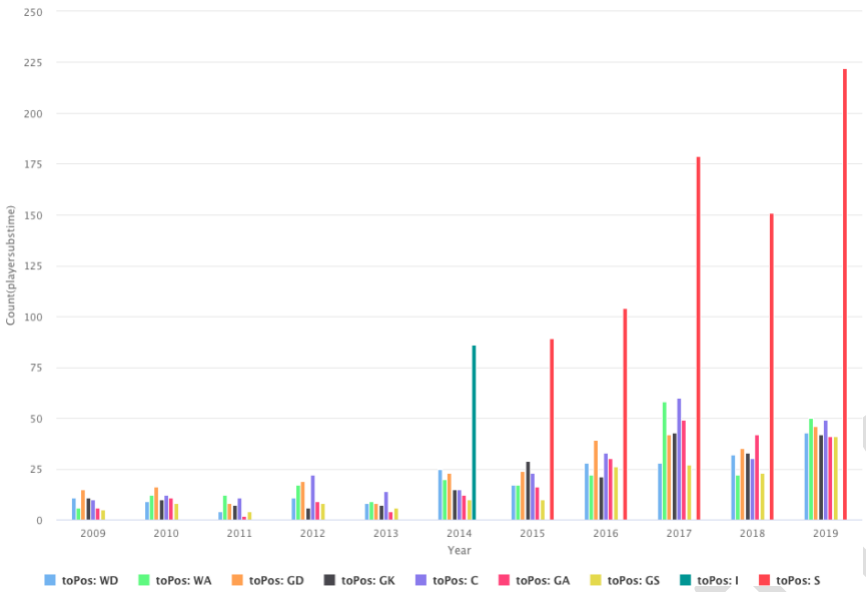


Fig 20: Bar chart for player substitution trend

- Based on bar chart, it was observed that most of the teams started exploiting the rule of interchangeability of players after 2014 with most of the substitutions are occurring to position "S" in every game.

3.3.2 Hypothesis 6: There are very less teams who are substituting players with the same position they are playing in the game.

<Substitution of players based on game position>

fromPos	Max. squadNa..	toPos GA	WA	GS	C	GK	GD	WD	I	S
GA	Tactix				1					
	Thunderbirds		15				1			
	Vixens								11	
	WBOP Magic			67						125
WA	Pulse							3		
	Tactix		2							
	Thunderbirds	14		2					18	
	Vixens									110
	WBOP Magic				112					
GS	Thunderbirds		3				2			
	Vixens			1						103
	WBOP Magic	76							10	
C	Thunderbirds	1								
	Vixens								15	124
	WBOP Magic		112		3			52		
GK	Thunderbirds			2		1			11	
	Vixens									83
	WBOP Magic						116	9		
GD	Thunderbirds	1					2			
	Vixens									139
	WBOP Magic					81		54	25	
WD	Mystics							3		
	Thunderbirds		6							
	WBOP Magic				86	11	49		17	91
I	Thunderbirds					16				
	Vixens	12	14		15					
	WBOP Magic			9			21	20		
S	Vixens		107	108		102		104		
	WBOP Magic	119			107		128			

1 139

Sum of Number of Records (color) broken down by toPos vs. fromPos and maximum of squadName.

Fig 21: Distribution of player substation according to position

Above exploratory analysis gives insight about how coaches are substituting players based on their "positions" in game. To make more advantage of interchanging player it can be seen as a trend where during winning coaches are making more player available in defense region and while losing a game, they are making more players in attacking region of the game.

On further evaluation, coaches are primarily doing substitution in game for changing position in game for players. To prove our hypothesis, predictive model have been created using rapid miner where below are the dataset, labels and predictors used for modelling -

Datasets – Teams and Playersubs

Label - topos

Predictors – frompos, periodseconds and period

Examples that are irrelevant in the data have been filtered and model has been created using cross validation for observing better result for data. Various model had been tried but Deep learning with 30 epochs was giving the best accuracy based on our dataset as shown in below figure.

New data has been created for observing result from deployment and hence the validated result from cross validation was used in apply model for observing best predicted results for our hypothesis.

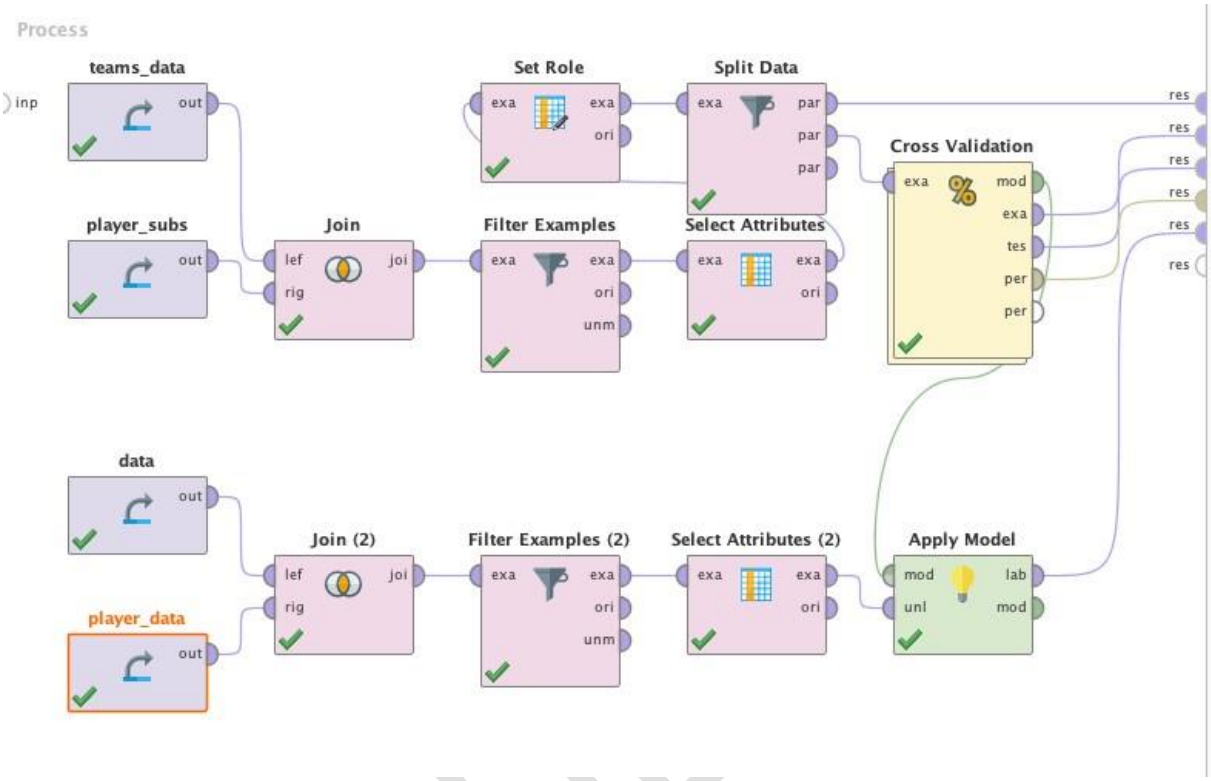


Fig 22: Model for Hypothesis 6

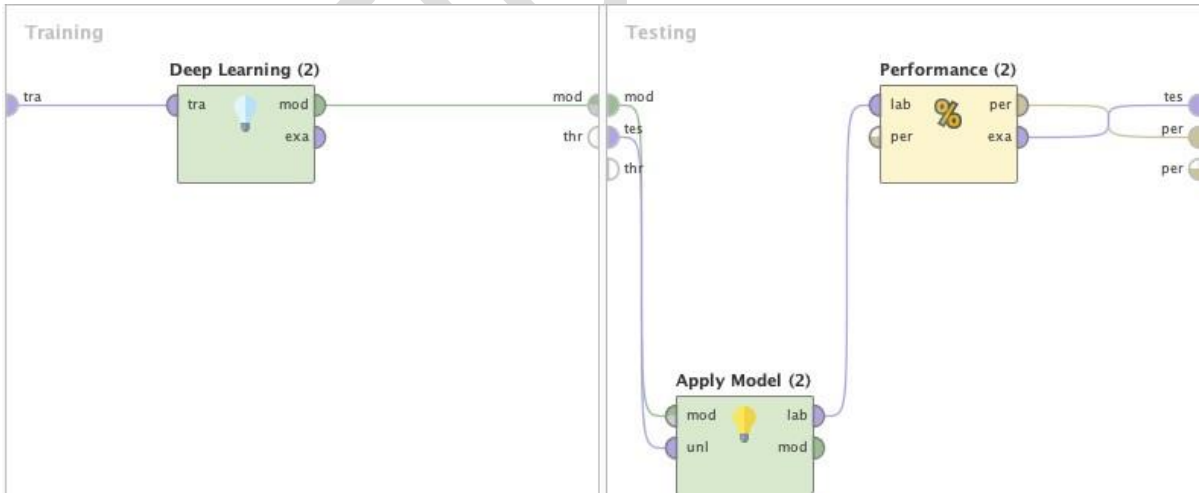


Fig 23: Cross validation sub process for hypothesis 6

accuracy: 44.39% +/- 2.01% (micro average: 44.39%)

	true WD	true GD	true WA	true C	true GK	true GA	true GS	true I	true S	class preci...
pred. WD	38	12	25	18	20	33	25	0	3	21.84%
pred. GD	30	99	19	23	24	29	16	11	28	35.48%
pred. WA	58	37	82	19	30	28	38	12	26	24.85%
pred. C	18	41	17	125	24	23	16	13	76	35.41%
pred. GK	83	67	68	65	110	82	59	6	28	19.37%
pred. GA	37	34	29	40	31	85	19	4	16	28.81%
pred. GS	12	11	18	10	7	13	41	5	20	29.93%
pred. I	0	1	0	0	0	0	0	1	0	50.00%
pred. S	60	86	85	95	61	52	50	44	1087	67.10%
class recall	11.31%	25.52%	23.91%	31.65%	35.83%	24.64%	15.53%	1.04%	84.66%	

Fig 24: Confusion Matrix from model

For same position the correct predicted results are very less compared to result for different positions proving our hypothesis. For instance, True and Predicted value for Wing Defense are only 21% for the whole model whereas player substituted from Wing defense to other positions covers 79% for the players playing in Wing defense position.

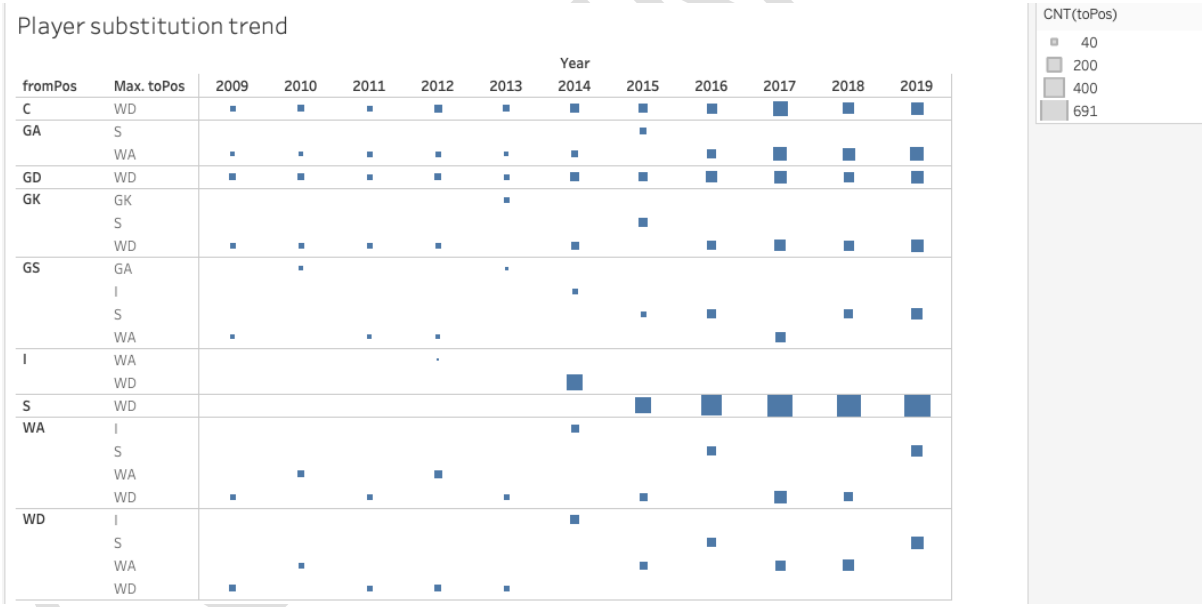


Fig 25: Player Substitution Trend

- It was observed that most of the teams start using substitution rule after 2014.
- Most of the times player are substituting from Centre position to Defence position and from wing attacking position to wing defence position and vice versa.
- Moreover in domestic league approx. 1700 times once player got substituted never got chance to play again in the game as their position changes to "S" after substitution.
- Overall "**Thunderbirds**" did maximum substitution from Goal attacker to Goal shooter whereas "**WBOP magic**" doing most of substitution to center forward from attacking and wing defense tactics during their game.

4. Recommendation

Based on analysis conducted, following recommendation are suggested:

Since players performance in international matches has reduced in compared to the domestic match, selection of players based on only SSN is not recommended. Players selection could be done based on overall performances, fitness and other attributes. Also, it should be noted that same level of performance should not be expected in domestic and international matches as external factors impact on performance of players in international matches.

Similarly, after the spilt into separate domestic league, performance have seen to be improved for Australian Diamonds. So, they should continue easing the restriction to bring new players of other countries to improve the performance of SSN.

Correct rotation/substitution of players during game and changing the position of players after every quarter can create more chances of winning the game. For instance, a player playing in centre position in first half can change their position to defence or attacking in second half.

Teams like Vixens, Giants who are top level teams in domestic competition and winning most of the matches have to use rule of substitution in a correct way to keep their players fit throughout the competition. helping teams to keep fitness level up-to data for all players and assisting in identifying new players that can contribute to their team.

Teams must use rolling substitutions to ensure equal playing time. Players can be swapped at any time during the game, with players meeting at sideline and tagging, so one player comes off and another goes on to boost up confidence of all players of the team.

There might be insignificant drop-off in performance in the middle of each quarter of matches, but team's best player can contribute better because they are more rested. Substituting best players at right time can make more chances of team winning in the game.

Player "interchanging" rule can be introduced in SSN as most of the teams are using rule of "player on the court" for interchanging the players which proves our hypothesis.

5. General Analytics Issue

Encountering ethical and analytical issues in analysis process is natural. Thus, some of the general analytical issues faced during this exploratory and advance analysis are as follows:

- **Data Quality Issues**: Since all data in given dataset were not structured thus, the findings and predictions can be misleading sometime. (Wang, Kon and Madnick, 1993).
- **Missing Values Issues:** Null and missing values in data were tackled through the assumption. However, it may result to biasness in predictions and results.
- **Human Error:** Since the analysis process involves the human intervention thus assurance of 100% accurate prediction is not possible (Tripathy, 2013).
- **Privacy and Ethical Issues:** As per the signed contract, the disclosure of data has been the main concern. In case of failure to so, would lead to breach of contract. Thus, data and information must

be shared on the bias of privacy norms and be altered to present the unanimous findings (GS1 Australia Privacy Policy - GS1 Australia, 2020).

- **Misinterpretation of data:** The deep learning and analysis of the data elements is key of understanding data otherwise the relevant and accurate insight is not possible (Keim, Lee, Thuraisinghaman, and Wittenbrink, 1995).

6. References

Gs1au.org. 2020. *GS1 Australia Privacy Policy - GS1 Australia*. [online] Available at: <https://www.gs1au.org/gs1-australia-privacy-policy/> [Accessed 12 April 2020].

Keim, D.A., Lee, J.P., Thuraisinghaman, B. and Wittenbrink, C., 1995, October. Database issues for data visualization: supporting interactive database exploration. In *Workshop on Database Issues for Data Visualization* (pp. 12-25). Springer, Berlin, Heidelberg.

Tripathy, J.P., 2013. Secondary data analysis: Ethical issues and challenges. *Iranian journal of public health*, *42*(12), p.1478.

Wang, R.Y., Kon, H.B. and Madnick, S.E., 1993, April. Data quality requirements analysis and modeling. In *Proceedings of IEEE 9th International Conference on Data Engineering* (pp. 670-677). IEEE.

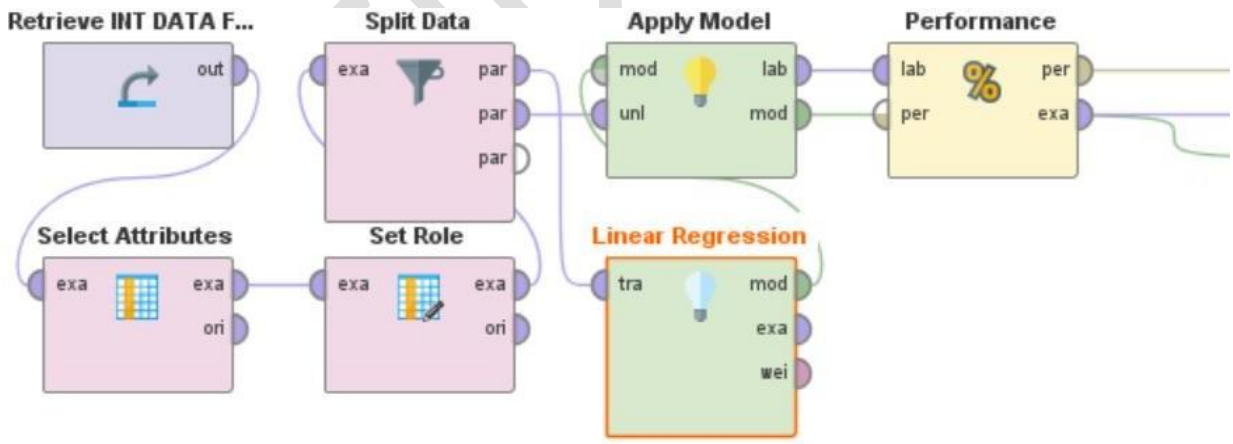
7. Appendix A

Hypothesis 1: Hypothesis Testing template

Goal attempts			
Hypothesis Test for μ (Mean)			
Hypotheses			
Null Hypothesis	μ	=	9.52631579
Alternative Hypothesis	μ	<	**9.52631579**
Test Type			left
Level of significance			
		α	**0.05**
Critical Region			
Critical Value (s)			-1.6449
Population Standard Deviation			11.6239649
Sample Data			
Sample Mean			**5.36756757**
Sample Size			**370**
Standard Error of the Mean			0.6043
Z Sample Statistic			-6.8819
p-value			0.0000
Decision			
Reject Null Hypothesis			

Hypothesis 2:

Predictive model using data of international rounds



Example dataset for Predictive model using data of international rounds:

Row No.	goalAttempts	predictio... ↓	goals	playerId
136	47	52.295	47	80113
80	50	51.184	46	80113
98	48	48.962	44	80113
147	46	46.740	42	80113
32	43	43.407	39	80113
125	40	42.295	38	80113
20	36	37.851	34	80052
106	40	35.629	32	80113
51	31	33.407	30	80113
83	32	33.407	30	993463
17	27	25.629	23	80072
26	24	24.518	22	80113

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.313 +/- 0.000
absolute_error: 0.630 +/- 1.153
relative_error: 12.89% +/- 10.26%
correlation: 0.994
squared_correlation: 0.987
```

Hypothesis 4: Performance vectors with different labels

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.125 +/- 0.000
absolute_error: 0.909 +/- 0.664
normalized_absolute_error: 1.038
root_relative_squared_error: 1.007
squared_error: 1.266 +/- 1.865
prediction_average: 3.500 +/- 1.118
```

Bad Hand

```
PerformanceVector:
root_mean_squared_error: 1.350 +/- 0.000
absolute_error: 1.050 +/- 0.848
normalized_absolute_error: 0.933
root_relative_squared_error: 1.020
squared_error: 1.822 +/- 2.391
prediction_average: 4.500 +/- 1.323
```

Bad Passes

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.501 +/- 0.000
absolute_error: 0.501 +/- 0.008
normalized_absolute_error: 1.002
root_relative_squared_error: 1.002
squared_error: 0.251 +/- 0.008
prediction_average: 0.500 +/- 0.500
```

Breaks

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 2.765 +/- 0.000
absolute_error: 2.186 +/- 1.692
normalized_absolute_error: 1.029
root_relative_squared_error: 1.011
squared_error: 7.645 +/- 8.374
prediction_average: 7.625 +/- 2.736
```

Goal Missed

```
PerformanceVector:
root_mean_squared_error: 15.526 +/- 0.000
absolute_error: 9.207 +/- 12.502
normalized_absolute_error: 0.893
root_relative_squared_error: 1.003
squared_error: 241.059 +/- 573.139
prediction_average: 24.875 +/- 15.479
```

Turnover

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.818 +/- 0.00
absolute_error: 1.612 +/- 0.842
normalized_absolute_error: 1.052
root_relative_squared_error: 1.011
squared_error: 3.307 +/- 2.797
prediction_average: 1.625 +/- 1.798
```

Offsides

Hypothesis 5&6:

Below table shows how many players are substituted based on “topos”

Substituted position	Number of substitutions
To position Centre(C)	1485
To position Goal attack (GA)	1400
To position Goal defense (GD)	1523
To position Goal keeper (GK)	1233
To position Goal shooter (GS)	910
To position Wing attacker (WA)	1401
To position Wing defense	1463
To position I	310
To position S	1463

Predicted results for Hypothesis 5 –

periodS... ↓	prediction(periodSeconds)	playersubstime	fromPos	toPos
919	422.028	[735.0 - 919.0]	WA	S
919	410.735	[735.0 - 919.0]	GK	S
914	286.388	[735.0 - 919.0]	C	S
904	316.422	[735.0 - 919.0]	I	GK
900	410.861	[735.0 - 919.0]	GD	GK
900	411.720	[735.0 - 919.0]	GS	S
900	247.105	[735.0 - 919.0]	GK	S
900	336.755	[735.0 - 919.0]	S	GK
900	407.445	[735.0 - 919.0]	GD	S
898	225.265	[735.0 - 919.0]	GA	S
898	210.819	[735.0 - 919.0]	S	WA
898	251.265	[735.0 - 919.0]	WA	S
898	250.688	[735.0 - 919.0]	S	GA
895	353.454	[735.0 - 919.0]	GK	S
894	336.659	[735.0 - 919.0]	I	GA
887	207.345	[735.0 - 919.0]	GK	S
886	323.032	[735.0 - 919.0]	GS	S
881	204.712	[735.0 - 919.0]	GD	GD
881	242.041	[735.0 - 919.0]	GK	S

Performance Matrix for Hypothesis 6 –

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 49.436 +/- 0.000
absolute_error: 42.897 +/- 24.572
relative_error: 17.33% +/- 109.99%
relative_error_lenient: 53.20% +/- 45.67%
squared_error: 2443.936 +/- 3164.011
correlation: 0.991
```

Predicted result for Hypothesis 6 –

toPos	prediction(...
GK	GK
GD	C
C	C
GA	GA
WD	GD
C	C
GD	GD
GA	GS
WA	C
GK	GK
C	C
WA	C
WD	C
GD	GD
WA	GS
C	C
GS	GS
GK	GK