International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

ANALYSIS, PREDICTIONS AND STRATEGY FROM KABADDI AN QUANTITATIVE APPROACH

Nunavath Rajitha
rajithanunavath123@gmail.com
Physical Director, Alphoras Degree PG College, Karimnagar , Telangana.

ABSTRACT:

Kabaddi is a contact team sport of Indian-origin. It is a highly strategic game and generates a significant amount of data due to its rules. However, data generated from kabaddi tournaments has so far been unused, and coaches and players rely heavily on intuitions to make decisions and craft strategies. This paper provides a quantitative approach to the game of kabaddi. The research derives outlook from an analysis performed on data from the 3rd Standard-style Kabaddi World Cup 2016, organised by the International Kabaddi Federation. The dataset, which consists of 66 entries over 31 variables from 33 matches, was manually curated. This paper discusses and provides a quantitative perspective on traditional strategies and conceptions related to the game of kabaddi such as attack and defence strategies. Multiple hypotheses are built and validated using studentâĂŹs t-test. This paper further provides a quantitative approach to profile an entire tournament to gain a general understanding of the strengths of various teams. Additionally, team-specific profiling, through hypotheses testing and visualisation, is presented to gain a deeper understanding of teamâĂŹs behaviour and performance. This paper also provides multiple models to forecast the winner. The model-building includes automatic feature selection techniques and variable importance analysis techniques. Generalised linear model with and without an elastic net, recursive partitioning and regression tree, conditional inference tree, random forest, support vector machine (linear and radial) and neural network-based models are built and presented. Ensemble models use generalised linear model and random forest model techniques as ensemble method to combine outcome of a generalised linear model with the elastic net, random forest, and neural network-based models. The research discusses the comparison between models and their performance parameters. Research also suggests that ensemble technique is not able to boost up accuracy. Models achieve 91.67% − 100% accuracy on cross-validation dataset and 78.57% − 100% on test set. Results presented can be used to design in-game real-time winning predictions to improve decision-making. Results presented can be used to design agent and environments to train artificial intelligence via reinforced learning model.

KEYWORDS
Kabaddi, Sports Analytics, Predictive Model, Team Profiling, Visualisation, Hypotheses, Model selection

INTRODUCTION

Kabaddi is a team contact sport which has originated in India and has travelled to other countries in the region. Kabaddi's uniqueness is that an entire team defends against a single attacking player from the opposition team. Appendix A provides further details about game of kabaddi. A majority of Indians either knows or plays kabaddi [28]. However, the use of technology to better the game is non-existent so far.

Commercialisation of technology has led to the penetration of the same into kabaddi as well [1]. Kabaddi produces data which is under-utilised today, and in the best form used for showingdescriptive statistics. Traditionally, intuitive judgements drove decision-making in kabaddi. This research aims at converting "*what we think*" to "*what we know*", an approach famously prescribed by Sam Allardyce [8]. The research investigates established claims and strategies of kabaddi and validates them using hypothesis testing methodology. This validation starts at the tournament level, and team level granularity is introduced as a next step. Based on the proven foundation, this research then ventures into predictive modelling of the game outcome using supervised learning method. Ensembling technique is used to make predictive models robust, though results are not promising. Results from predictive models are highly satisfactory with 100% prediction accuraciesdemonstrated.

Literature survey, research method, dataset preparation, descriptive statistics, hypotheses validation, model-building and discussion of results, form the seven sections of the paper. Literature survey discusses the state-of-the-art for kabaddi sports analytics. Research method and dataset preparation discuss the data generation and curation along with research method used. Descriptivestatistics section provides a detailed understanding of dataset via numbers and visualisations. Hypothesis validation investigates common strategies and claims of kabaddi. Model building forms the core of research and provides adetailed discussion of predictive models andtheir accuracy. Discussion of results comprises of the prospects of research findings along with the possible applications.

LITERATURE SURVEY:

tached to each sport [15]. India iswitnessing the rise of professional sports, banking on the success of various leagues [31]. Sports is all aboutdecision-making on the field and off the field, considering multiple parameters. Kabaddi, as a sport, is not different in this respect. Kabaddi can benefit from analytics as it produces a variety of data at the team level and individual player level. Devenport [8], in his comprehensive article, recommends three types of analytics for sports - players' injury and health analytics to predict their fitness and readiness for a game, business analytics to leverage business aspects of a game, and player and game performance analytics to help predict individual and game outcomes.

Player's injury and health indicators are measured meticulously, and various studies leverage this information. The research by De, Dasgupta, Panda, and Bhattacharya [9] is one of the early efforts to test the physical efficiency of a male player and relate it to their health. The research by Khanna, Majumdar, Malik, Vrinda, and Mandal [19] is another study to measure physiological responses of players during the game to deepen understanding of player health.

Additionally, specific injuries like knee injuries [11] of players and most common injuries [22] are subjects of research. These studies also provide probable causes for injuries. There is a need to bring fragmented and piecemeal research on players' injury and health together to consolidate the learnings and present a holistic health and injury analytics for kabaddi players, coaches, and team managers. Business analytics, which includes fan programmes and engagement, dynamic ticket pricing, and marketing optimisation, is not into the exclusive focus of researchers. The research by Sanjeev and Ankur [31] on the topic of "Constituents of Successful Sports Leagues in Emerging Markets," discusses aspects of fan engagement and celebrity endorsement for kabaddi at the surface level.

In the game of Kabaddi, player performance is a crucial indicator of the success of the team. Studies have correlated player performance based on fitness [27] , cognitive abilities [18] [30]. However,playerperformancedetermination remains an area for further research. Additionally, detailed literature survey highlights that no substantial research is available concerning game performance and game outcome predictions. Though a method to record and perform analysis of kabaddi matches are proposed [24], it did not attract further attention.

Analytics is visibly reshaping how sports are played [13]. Analytics helps players in understanding their weaknesses and strengths, assists the coaches in making informed decisions rather than intuitive decisions, and helps managers optimise the costs. Sports analytics is thriving today due to the demonstration of its impact by practitioners and groups who believed in it. Sports analytics is becoming the new standard and is continuously evolving the operational and strategical aspects of the game. Literature review, however, points out, in general, the lack of application of sports analytics in the game of kabaddi. The motivation of this research is to demonstrate that sports analytics can be applied to games like kabaddi. This research focuses on the aspects of game performanceanalytics. Proprietary sports statistics may not help sports analytics movement in general [13] and this is another motivation for publishing research on kabaddi sports analytics.

METHODLOGY:

3 RESEARCH METHOD

The objective of this research is to build a model to predict the outcome of the game while validating the established claims and strategies. Research methodology consists of curating the dataset and performing a variety of analytics on it to provide insights.

The dataset is processed to produce a descriptive statistical summary along with visualisation to improve the understanding of the dataset. Multiple hypotheses are validated using parametric hypothesis test. The predictive model building includesderiving multiple models and later choosing them based on performance parameters. The research method applied here is the most common method employed by analytics practitioners.

4 DATASET PREPARATION

Data and dataset preparation is critical as it forms the core of any analytics exercise. Dataset is prepared by collecting raw information and then applying processing techniques to make the data relevant for research. For this research, data isconsolidated by manually scraping through the website of 3rd Standard-styleKabaddi World Cup tournament. The postprocessed dataset consists of 66 entries over 31 variables from 33 matches. Curated dataset along with its codebook, which describes variables and its properties, are published on Kaggle platform [25]. Appendix B provides more information about dataset and codebook. This curated dataset has received '*featured*' status on Kaggle platform as it is the first dataset published on kabaddi. Variables of the dataset represent the attack and defence points acquired by teams as per standard kabaddi rules like tackle points, raid points, bonus points, allout points, touch points, total points and other points. Variables representing relative differences in points for tackle, touch, raid, bonus, total, and alloutmanoeuvres are calculated in apostprocessing phase. The dataset

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST

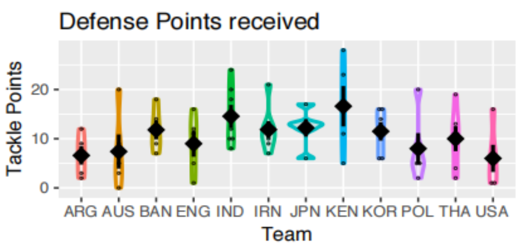
also captures results of a game, results oftoss and game stage of the league. This dataset is complete and consistent to sharpen further analytics.

5 DESCRIPTIVE ANALYSIS

Descriptive analytics is the primitive form of analytics which utilises past data to provide insights. Descriptive statistics and data visualisation are fundamental elements of descriptive analytics [21] that provide an innovative way to summarise data.Descriptive statistics of location measures (mean, median) and dispersion measures (standard deviation, mean absolute deviation, range, skew, kurtosis, standard error, and interquartile range) of the essential variables is available in Table 1. Variables capturing relative difference have mean, median and skew parameters as zero due to the symmetry. The observed distribution of all variables is platykurtic with a positive skew. The summary provides surface level insights: every team observes one allout per match on an average, two super tackle per three matches, and likewise. Additionally, a team receives 18 touch points, 5 bonus points, 11 tackle points, 4 allout points, and 2 extra points in a game totalling to 41 points. Team-wise information is necessary to deepen the understanding

Table 1: Descriptive statistics of essential variables

Variables	n	mean	sd	median	mad	min	max	range	skew	kurtosis	se	IQR
alloutRec	66	2.03	2.14	1	1.48	0	7	7	0.69	-0.95	0.26	4
sTackleRec	66	0.65	1.36	0	0	0	9	9	3.78	18.88	0.17	1
touchPntsRec	66	17.85	9.17	15.5	9.64	1	37	36	0.36	-0.88	1.13	14.75
bonusPntsRec	66	5.47	3.03	5	2.97	0	16	16	0.81	0.9	0.37	3.75
raidPntsRec	66	23.32	10.58	20.5	11.12	1	43	42	0.33	-0.96	1.3	16.5
tacklePntsRec	66	10.64	6.24	11	7.41	0	28	28	0.44	-0.3	0.77	8
alloutPntsRec	66	4.06	4.28	2	2.97	0	14	14	0.69	-0.95	0.53	8
extraPntsRec	66	2.45	2.01	2	1.48	0	8	8	0.75	-0.37	0.25	3
totalPntsRec	66	40.47	19.49	34	20.76	8	80	72	0.45	-1.17	2.4	30.5
touchPntsDiff	66	0	15.18	0	19.27	-33	33	66	0	-0.93	1.87	25.5
bonusPntsDiff	66	0	3.79	0	4.45	-9	9	18	0	-0.22	0.47	6
raidPntsDiff	66	0	17.32	0	20.76	-42	42	84	0	-0.67	2.13	27.5
tacklePntsDiff	66	0	10.37	0	11.86	-21	21	42	0	-1	1.28	15.5
alloutPntsDiff	66	0	7.86	0	11.86	-14	14	28	0	-1.31	0.97	16
extraPntsDiff	66	0	3	0	2.97	-7	7	14	0	-0.29	0.37	4
totalPntsDiff	66	0	35.66	0	50.41	-72	72	144	0	-1.19	4.39	68



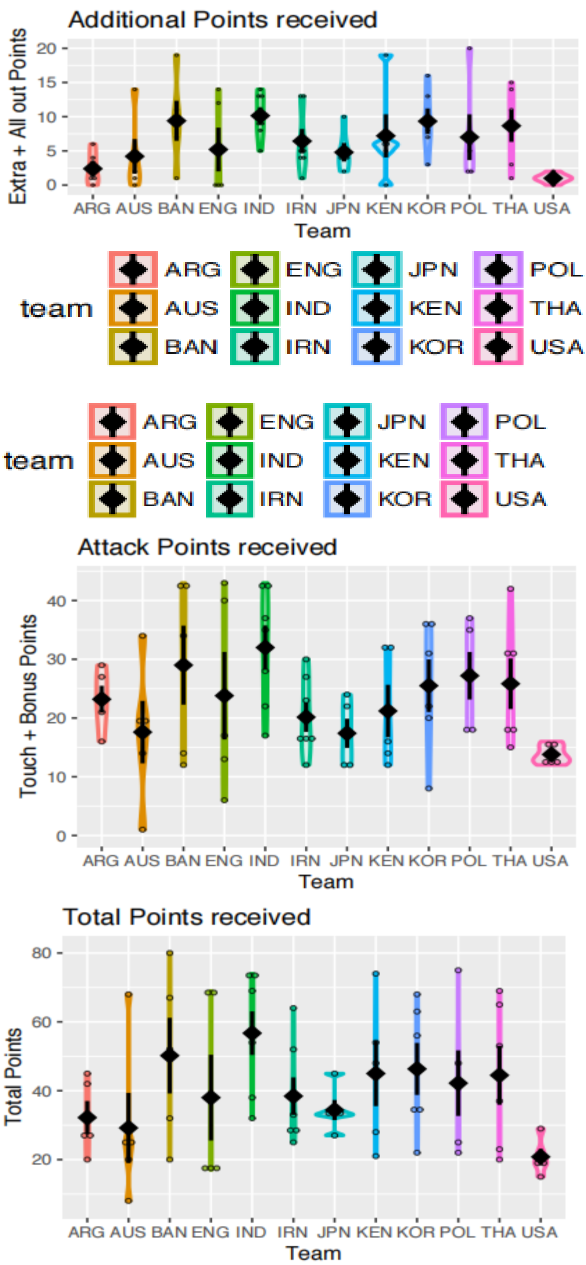
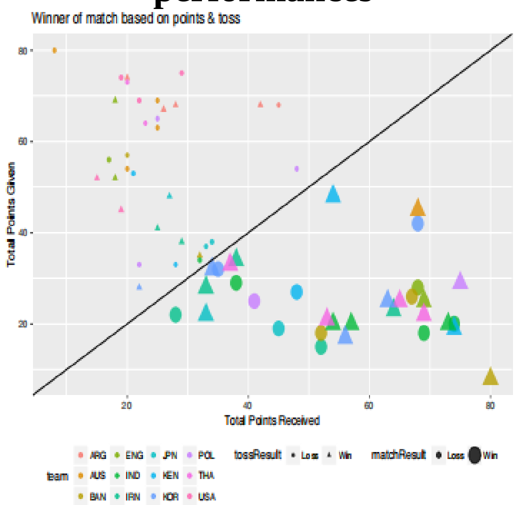


Figure 1: Violin plots of team performances



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

Figure 2: Results of match outcome based on total points andtoss results

of team performance. Team details are shown using violin plots in Figure 1, which are built using box plots and density plots to provide complete information about the given variable [17]. Visualisation is a powerful tool to draw insights. Analysis of Figure 1 presents some critical insights like Kenya has the best defence among the teams and India has the best attack. Based on the average of total points, India stands as the best team followed by Bangladesh and South Korea. Figure 2 provides match results considering to tal points and toss results. Analysis of Figure 2 demonstrates that there is no impact of toss result on the winning team and the total points scored, hence rendering toss results useless in deciding the game outcome. Descriptive analytics provides necessary insights into the performance of the game at the aggregate level and at the team level. It also helps in building predictive models by providing dependent variable characteristics.

6 HYPOTHESES VALIDATION

Humans are efficient in finding the patterns and codifying that knowledge into thumb rules. The process of learning is lengthy and slow and is subject to cognitive biases. Majority of the decisions of kabaddi game, on the field and off the field, are based on traditional wisdom passed on as thumb rules from generations of coaches and players. This research aims at investigating some of the known thumb rules and validating them using the statistical framework of hypothesis testing. Hypothesis testing is performed using student's t-test for the small dataset [10].

Discussion with coaches, players, and journalists points to four thumb rules often used in decision making.

- Attack is better than defence
- Defence is better than extra
- Defence is better than allout
- Allout is better than extra

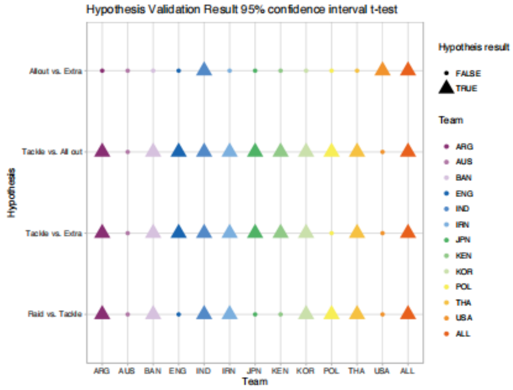


Figure 3: Team-wise hypothesis validation results with 95%confidence interval

For validation of thumb rules, research derives hypotheses, and testing results of the same are available in Table 2. Hypotheses testing uses single-sided (higher than) student's t-test with a paired sample having unknown variance. It is clear from the results of hypotheses testing that there is no statistical evidence to refute the thumb rules even with the stringent confidence interval of 99.99%. Based on the results, decision making can be generalised in a simple way using priority order. Priority ofattack is highest followed by defence, allout and extra. This statistically proven rule is helpful in deciding strategies to be implemented.

This hypothesis testing is extended to understand team wise application of thumb rules. Results of hypothesis testing for each team along with aggregate are presented graphically in Figure 3 with 95% confidence interval. The observation of Figure 3 yields that Argentina, Bangladesh, India, Iran, Korea, Poland, and Thailand follow the thumb rule of "attack is better than the defence." However, the same thumb rule might not befollowed by Australia, England, Japan,Kenya, and the USA. Interestingly, Indiafollows all thumb

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST ISSN: 2457-0362

rules, but Australia follows none. Hypothesis visualisation is a powerful technique, to understand the working mechanism and deeply-held beliefs of various teams, which helps in developing a game plan and strategy.

7 MODEL-BUILDING

Models predict the binary game outcome as either loss or win by processing selected features using supervised learning method.

Model-building is an iterative process which considers robustness and accuracy improvement as the prime measures during the development phase. Before kick-starting the model building process, research divides data into three parts. The dataset is split into ˜60% (40 samples) of training data to train and build a model, ˜20% (12 samples) of cross-validation data to improve and tune the model, and ˜20% (14 samples) of test data to check the final accuracy and robustness of the model. Understanding of variables and their features in the

Table 2: Four key hypotheses validation results

Thumb Rule	Null Hypothesis	Alternate Hypothesis	p-Value	Result
Attack is better than defence	Raid points of winning team is same as defence points	Raid points of winning team is higher than defence points	8.28E-11	Null Hypothesis Rejected with >99.99% of confidence interval
Defence is better than extra	Defence points of winning team is same as extra points	Defence points of winning team is higher than extra points	3.91E-14	Null Hypothesis Rejected with >99.99% of confidence interval
Defence is better than allout	Defence points of winning team is same as allout points	Defence points of winning team is higher than allout points	5.19E-11	Null Hypothesis Rejected with >99.99% of confidence interval
Allout is better than extra	Allout points of winning team is same as extra points	Allout points of winning team is higher than extra points	7.41E-08	Null Hypothesis Rejected with >99.99% of confidence interval

dataset is an essential first task in building models, widely known as feature engineering. Feature engineering helps in identifying predictor variables to improve the performance of the models. It is a demanding process involving a profound understanding of data and the domain. Feature engineering method can be automated in a limited way with the help of RFE - recursive feature elimination algorithm [29] which uses backward feature selection process [16].

RFE algorithm is configured to use the random forest as a helper function along with five times repeated cross-validation over 20 random samples. Research restricts the set of input features for RFE to 21 variables excluding total points, total points difference, match stage, toss results, and count of allouts. Total points and total point difference are apparent predictors of game outcome and can intro duce heavy bias. Match stage and toss results have no impact on the game outcome. Allout points are representative of allout counts.

RFE, from the available set of variables, provides the optimal set of 8 variables as best fit. However, only top 5 predictors - *touchPntsDiff*, *raidPntsDiff*, *alloutPntsDiff*, *tacklePntsDiff*, and *alloutPntsRec* - are used to build various models.

In total, ten base models are built using logistic regression-based algorithms, tree-based algorithms, support vector machine-based algorithms and neural network-based algorithms. Logistic regression based models include generalised linear model (*glm*), and a generalised linear model with an elastic net (*glmnet*). Tree-based models include recursive partitioning and regression trees (*rpart*), conditional inference tree (*ctree*), and random forest (*rf*). Support vector machine based models include support vector machine with linear kernel (two types of implementation; *svmLinear*, *svmLin ear2*), regularised support vector machine with a dual linear kernel (*svmLinear3*), and support vector machine with radial basis kernel function (*svmRadial*). Neural network-based model includes a single hidden layer feed-forward neural network (*nnet*). Pure models performances regarding accuracy, sensitivity, specificity and F1 score are available in Table 3 for cross-validation dataset and test dataset [26]

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

[7]. From Table 3, it is clear that *rpart* and *rf* exhibit overfitting. The accuracy of *ctree* is the lowest but does not display symptoms of overfitting. All remaining models provide accuracy of 100% over cross-validation dataset and test dataset.

Despite similar performances, it is necessary to understand the decision-making process of these models. The importance of vari ables in the context of the model provides the necessary under standing of decision-making process [14] [23] and Table 4 depicts it. The accuracy of *ctree* is lowest despite the same importance of variables for support vector machine-based models. The difference in processing of variables by the respective algorithm is the primary cause of different accuracies. The *glm* and *glmnet* have an entirely different order of importance despite using same logistic regressionbased techniques. It is due to the presence of the regularisation mechanism in *glmnet*. It is clear that various models treat variables differently and hence there is a possibility of overfitting for a given model, which is difficult to discover in this small dataset.

Ensemble mechanism is a technique used to combine multiple models, referred as base learners, to improve the overall performance of predictions [12] [20] with increased robustness. Variation in the importance of variables for a set of base learning models is a dominant criterion used for selection of base learners. These base learners are combined using *rf* and *glm* as ensemble operator. Base learners considered for ensemble are *rf*, *nnet*, and *glmnet*. Ensemble operator *rf* favours base learner *glmnet* (100%), followed by *rf* (42.21%), and discards *nnet* (0%) entirely in the decision-making process. Ensemble operator *glm* behaves in complete contrast to that of *rf* and heavily favours base learner *nnet* (100%), followed by *glmnet* (3.36%), and ends up discarding *rf* (0%) entirely. Ensemble models performance is same as that of their heavily favoured base learners and hence failed to demonstrate any performance boost. Ensemble method with *rf* operator is preferred, as it reflects the outcome of at least two models in the decision-making process.

Pure model *nnet* is preferred, as its importance of variable matches the results of hypotheses validation.

8 DISCUSSION OF RESULTS

The results of visualisation, hypothesis testing, and predictive model building are quite encouraging. Before generalising the interpreta tion of results, caution needs to be exercised as results can suffer from extreme biases induced by small sample size. Despite this limitation, techniques and methods used for research can be generalised.

The violin plot discussed as part of descriptive analytics is a useful tool to visualise the behaviour of teams comprehensively.

Granular information provided in the violin plot can be used to characterise a team which is particularly helpful in devising winning strategies.

Hypothesis testing results reinforce common prevailing wisdom and provide statistical backing for the same. It also highlights an important but expected aspect that, thumb rules of kabaddi reflect

Table 3: Performance measures (%) of models over cross-validation dataset and test dataset

		Cross-validation dataset (12 samples)				Test dataset (14 samples)			
		Accuracy	Sensitivity	Specificity	F1 Score	Accuracy	Sensitivity	Specificity	F1 Score
Models	*glm*	100	100	100	100	100	100	100	100
	glmnet	100	100	100	100	100	100	100	100
	rpart	100	100	100	100	78.57	100	57.14	82.35
	ctree	91.67	100	83.33	92.3	92.86	100	85.71	93.33
	rf	100	100	100	100	78.57	100	57.14	82.35
	svmLinear	100	100	100	100	100	100	100	100
	svmLinear2	100	100	100	100	100	100	100	100
	svmLinear3	100	100	100	100	100	100	100	100
	svmRadial	100	100	100	100	100	100	100	100
	nnet	100	100	100	100	100	100	100	100

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST

Table 4: Relative variable importance (%) for ten base learners

		Variables				
		touchPntsDiff	raidPntsDiff	alloutPntsDiff	tacklePntsDiff	alloutl
Models	glm	5.59	100	0	52.22	27
	rpart	49.49	0	73.54	1.8	1[illegible]
	ctree	100	79.31	86.21	0	10
	rf	100	58.94	49.8	9.23	[illegible]
	svmLinear	100	79.31	86.21	0	10
	svmLinear2	100	79.31	86.21	0	10
	svmLinear3	100	79.31	86.21	0	10
	svmRadial	100	79.31	86.21	0	10
	nnet	71.26	100	39.13	89.77	[illegible]
	glmnet	10.92	55.16	67.62	0	1[illegible]

that of the Indian team. At the aggregate level, hypotheses are withtanding even without considering Indian team's performance. The other teams demonstrate statistical backing for a maximum of three thumb rules. Visualisation technique demonstrated in hypothesis testing results can be combined with violin plots to understand and codify the real characteristics of any given team. Combined visualisation is beneficial for coach and teammanagement as it can assist in deciding strategies and line up for a game.

Predictive models, presented in research, are helpful in predicting the game outcome in real time with the available partial data.

Discussed models use only five simple dependent variables and provide high accuracy, making them a potential candidate to consider for deployment. Despite accuracies, models need to prove their validity and robustness in real life before deployment.

Ensemble models and neural network-based models provide superior accuracies, but they are difficult to interpret. Interpretability limits their application only to predict game outcomes and does

not help in devising strategies. For devising strategies, coefficients of glmnet are helpful. These coefficients provide the quantitative impact and correlation of each dependent variable on the game outcome. Coaches and players can easily interpret them and then devise strategies and scenarios.

Ensemble models and neural network-based models can help in testing the impact of strategies in various game scenarios. Scenariobased testing can assist the coach in training players for eventualities. Preparing all scenarios is a difficult task for coaches and players.

The solution to this problem exists in reinforced learning models.

In reinforced learning models, agents act on the environment in an

intelligent and in a self-learned way to maximise the given objective

over a period. Reinforced learning needs an environment which can be achieved by game modelling with standard kabaddi rules.

It also needs intelligent agents which can be modelled using discussed predictive models. In the process of learning, agents develop artificial intelligence to play and win the game. Agents can derive and devise optimal strategies for all scenarios. Artificial intelligence can open up a new dimension of unique game strategies. In an ideal world, reinforced learning mechanism can at least offset the role of a coach, if not replace, in deciding strategies.

The presented research and its findings have far-reaching implications for the game of kabaddi. The study offers a new and unique way to look kabaddi from a quantitative perspective. Thispaper paves the way for further research to transform the game of kabaddi using analytics.

CONCLUSION :

Kabaddi is considered a native Indian game devoid of any technologyled augmentation so far. The research aims at breaking this stereotype and provides evidence that kabaddi can benefit from applying analytics. Research method and its findings are cutting-edge and have the potential to change the future of the game. Detailed and dedicated research is necessary to deepen the understanding of theKABADDI: From an intuitive to an quantitative approach for analysis, predictions and strategy ,game and to extract further benefits. Predictive models combined with hypotheses and visualisation are the essential takeaways from the study.

A ABOUT GAME OF KABADDI

As per wikipedia [5], Kabaddi is a contact team sport. It is popular in South Asia and is the state game of the Indian states of Tamil Nadu, Kerala, Andhra Pradesh, Bihar, Haryana,Karnataka, Maharashtra, Punjab and Telangana.

A.1 Generic play of kabaddi

As per wikipedia [5] Kabaddi is played between two teams of seven players; the object of the game is for a single player on offence - referred to as a "raider" - to run into the opposing team's half of a court, tag out as many of their defenders as possible, and return to their own half of the court - all without being tackled by the defenders. Points are scored for each player tagged by the raider, while the opposing team earns a point for stopping the raider.

Players are taken out of the game if they are tagged or tackled, but can be"revived" for each point scored by their team from a tag or tackle.

A.2 Kabaddi world cup

The standard style Kabaddi World Cup, is an indoor international kabaddi competition conducted by theInternational Kabaddi Federation (IKF),contested by men's and women's national teams [6]. The competition has been previously contested in 2004, 2007 and 2016. All the tournaments have been won by India [6].

B KABADDI WORLDCUP DATASET

The 2016 Kabaddi World Cup, the third standard-style Kabaddi World Cup, was an international kabaddi tournament governed by the International Kabaddi Federation, contested from 7 to 22 October 2016 in Ahmedabad, India [4]. Twelve countries had competed in the tournament. 30 league matches played between teams. teams were deivided in 2 pools with 6 team in each pool. Top 2teams from each team were qualifid for semifinals and winner of semifianls played in finals.

This dataset contains data for all 33 matches at granualirity level of attack, defense, allout and extra points. Data set also includestoss results, super tackle count and all out count along with match results.

B.1 codebook

This dataset was manually prepared from taking necessary statistics from Kabaddi world cup site [2]. Points acquired as per rules are main statistics . This dataset contains necessary statistics in today format and details of all variables are as perfollowing.

• *gameNo* : Match number. Sequential Integer

• *team* : Team name Factor

• *oppTeam* : Opposition team name Factor

• *matchStage* : Tournament stage at which match was played.
(0 - League, 1 - SemiFinal, 2 - Final) Factor

• *tossResult* : Results of toss to select either side or raid (0 - Loss, 1 - Win) Factor

• *alloutRec* : No. of time team was all out yielding 2 point Integer

• *alloutGiv* : No. of time opposition team was all out yielding 2 point Integer

• *sTackleRec* : No. of times super tackle by team yielding 2 point Integer

• *sTackleGiv* : No. of times super tackle by opposition team yielding 2 point Integer

• *touchPntsRec* : No. of times player in raid touched opposition team player yiedling 1 point for every touch Integer

• *touchPntsGiv* : No. of times opposition player in raid touched team player yiedling 1 point for every touch Integer
• *bonusPntsRec* : No. of times player in raid crossed bonus line yiedling 1 point for every raid Integer
• *bonusPntsGiv* : No. of times opposition player in raid crossed bonus line yiedling 1 point for every raid Integer
• *raidPntsRec* : No. of total raid (attack) points by team, sum of touch points and bonus points Integer
• *raidPntsGiv* : No. of total raid (attack) points by opposition team, sum of touch points and bonus points Integer
• *tacklePntsRec* : No. of tackle (defense) points received by team yielding 1 point for normal tackle and 2 points for super tackle Integer
• *tacklePntsGiv* : No. of tackle (defense) points received by opposition team yielding 1 point for normal tackle and 2 points for super tackle Integer
• *alloutPntsRec* : No. of all out points received by team yielding 2 points per allout Integer
• *alloutPntsGiv* : No. of all out points received by opposition team yielding 2 points per allout Integer
• *extraPntsRec* : No. of extra (technical, penalty) points received by team Integer
• *extraPntsGiv* : No. of extra (technical, penalty) points received by opposition team Integer
• *totalPntsRec* : No. of total points received by team, sum of raid points, tackle points, allout points & extra points Integer
• *totalPntsGiv* : No. of total points received by opposition team, sum of raid points, tackle points, allout points & extra points Integer
• *touchPntsDiff* : No. of touch points difference from opposition team Integer
• *bonusPntsDiff* : No. of bonus points difference from opposition team Integer
• *raidPntsDiff* : No. of raid points difference from opposition team Integer
• *tacklePntsDiff* : No. of tackle points difference from opposition team Integer
• *alloutPntsDiff* : No. of allout points difference from opposition team Integer
• *extraPntsDiff* : No. of extra points difference from opposition team Integer
• *totalPntsDiff* : No. of total points difference from opposition team Integer
• *matchResults* : Results of the match (0 - Loss, 1 - Win) Factor

REFERENCES:

[1] 2014. Pro kabaddi league will encourage ancient sport. *The Hindu Business Line* (Jul 2014). http://www.thehindubusinessline.com/todays-paper/tp-others/ tp-states/'Pro-Kabaddi-League-will-encourage-ancientsport'/article20822336. ece
[2] 2016. Fixtures : 2016 Kabaddi World cup. (2016). http://www. 2016kabaddiworldcup.com/fixtures.html
[3] 2016. Rules : 2016 Kabaddi World cup. (2016). http://www.2016kabaddiworldcup. com/rules.html [4] 2018. 2016 Kabaddi World cup. (Feb 2018). https://en.wikipedia.org/wiki/2016_ Kabaddi_World_cup
[5] 2018. Kabaddi. (Jan 2018). https://en.wikipedia.org/wiki/Kabaddi
[6] 2018. Kabaddi World Cup (Standard style). (Feb 2018). https://en.wikipedia.org/ wiki/Kabaddi_World_Cup_%28Standard_style%29
[7] Rich Caruana and Alexandru Niculescu-Mizil. 2006. An Empirical Comparison of Supervised Learning Algorithms. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*. ACM, New York, NY, USA, 161–168.https://doi.org/10.1145/1143844.1143865

[8] Thomas Davenport. 2014. Analytics in sports: The new science of winning. *International Institute for Analytics* 2 (2014), 1–28.

Nunavath Rajitha :

I Nunavath Rajitha ,presently working as Physical Director in Alphores Women's Degree and PG college, karimnagar.I have completed MBA in sathavahana University in 2012. I have completed BPED in Barkatullah University in 2014.I have completed MPED in Vikrama simhapuri University, Nellore with 73% in 2018 . I have participated in many games in schools and college levels and got many medals.