Fuzzy Based Analysis of Proposed Model for Physical Health Standard Based on Association Rule Mining Techniques

Parneet Kaur[1], Dr Sawtantar Singh Khurmi[1], Dr Gurpreet Singh Josan[2]
[1]Sr Assistant Professor, Department of CSE, ACE & AR, Mithapur, Near Devsthali, Haryana, India.
[1]Head and Professor, Department of CSE &IT, BMEC, Muktsar, Punjab, India
[2]Professor, Department of CSE & IT, Punjabi University, Patiala, Punjab, India.

Abstract: **Data mining is used for mining data from databases and finding out meaningful patterns. In this paper authors have reviewed the literature of Data Mining based Association Rule techniques applied in the field of sports. This paper presents an approach for the analysis of proposed model, done on the Physical Health Standard data of students of particular age groups. The synthetic data is applied for the formation of fuzzy based association rules using MATLAB Mamdani Fuzzy Inference System. During this analysis, various parameters like Height, Weight, BMI, Speed Test, Balance Test, Flexibility Test, Medical Test, Endurance test are taken and membership functions are formed for deriving input and output attributes. As a result, lots of useful fuzzy based rules are generated for prediction of physical fitness. This analysis helps in the selection criteria and decision making system of various sports mining.**

Keywords: **Data Mining, PHS, FIS, Apriori, Association Mining Algorithm.**

I. INTRODUCTION

Data mining involves searching through databases for potentially useful information such as knowledge rules, patterns, regularities, and other trends hidden in the data. Numerous organizations in the field of business have shown that great success and lucrative outcomes can be accomplished through implementing data mining. For example, Wal-Mart used data mining and found a link between the sales of babies' diapers and beer. Based on this result, Wal-Mart placed beer close to the babies' diapers, which resulted in a significant increase in terms of beer sales [10]. Data Mining has emerged as a great tool for gathering useful information. Apart from algorithms, it makes use of various techniques like Fuzzy Logic, Neural Networks along with Decision trees, Nearest Neighbour method, Rule Induction and Apriori Algorithm. In recent years, the number and variety of applications of fuzzy logic haveincreased significantly. The applications range from consumer products such as cameras, camcorders, washing machines, and microwave ovens to Industrial process control, medical instrumentation,

Decision Support systems. Fuzzy logic has two different meanings. In a narrow sense, fuzzy logic is a logical system, which is an extension of multivalued logic. However, in a wider sense fuzzy logic (FL) is almost synonymous with the theory of fuzzy sets, a theory which relates to classes of objects with un-sharp boundaries in which membership is a matter of degree [13]. In this paper Fuzzy Inference System is used to derive fuzzy based rules for Student Physical Health Standard measurement. This Fuzzy when inter-disciplined with Data Mining is a relatively new field that brings together concepts from various parameters used in the work.

II. ASSOCIATION RULE MINING

Association rule mining is one of the most important and well researched techniques of data mining. An association rule is based on the form XY where X and Y are transactional item sets taken from a transactional database. Apriori is the most important algorithm. It has been developed for rule mining in large transactional database.

III. APRIORI ALGORITHM

Apriori Algorithm was proposed by Agarwal and Srikant in 1994. Apriori uses a complete, bottom-up search with a horizontal layout and enumerates all frequent item sets. The main property of Apriori algorithm is that all non-empty subsets of a frequent item set must also be frequent [10]. The original pseudocode by Agrawal and Srikant is given below:

Algorithm
Ck: Candidate item set of size k
Lk : frequent item set of size k
L1 = {frequent items};
for (k = 1; Lk != Uk; k++) do
begin Ck+1 = candidates generated from Lk; for each transaction t in database do increment the count of all candidates in Ck+1 that are contained in t . Lk+1 = candidates in Ck+1 with min_support end
return Uk Lk [12].

Association rule (AR) is commonly understood as an implication X → Y in a transaction database D = {t_1……t_m}.

Each transaction t_i Є D contains a subset of items I = $\{i_1 \ldots\ldots i_n\}$. X and Y are disjoint item sets, it holds X, Y belongs to I and X ∩ Y = Ø. The left hand side of this implication is called antecedent, the right hand side is referred to as consequent. The transaction database D can also be viewed as a boolean dataset where the boolean values of attributes in records express occurrence of items in transactions.

Support and confidence are the two most important quality measures for evaluating the interestingness of a rule.

IV. ASSOCIATION RULE BASED ALGORITHM

Many algorithms for generating association rules were presented over time. Some well known algorithms are: Apriori algorithm that uses breadth first search strategy, FP-Growth algorithm that uses an extended prefix tree structure to store the database in a compressed format. Cumulate and Estmerge algorithms to mine generalized association rules, QUARG to mine quantitative association rules, Extended Apriori to mine fuzzy generalized association rules, Apriori star to mine association rules in multiple tables and many more. Partition algorithm proposed by Ashok Savasere and others is a fundamentally different algorithm that reads the database almost two times to generate all significant association rules. Dynamic Hashing and Pruning Association Rule Mining using Hash Based Algorithm to filter the unnecessary items can be found in an effective hash based algorithm for mining association rules.

V. APPLIANCE OF ASSOCIATION RULE MINING ALGORITHMS IN SPORTS

Earlier Data Mining based algorithms were employed in business based applications but now much new research is going on to find and apply the same in finding knowledge for Physical Health standard measurements which further can be applied in sports. The advantage for sports organizations when it comes to data mining is in the resulting performance of their respective teams and players. Some sports are currently more advanced than others [13].The sports world is known for the vast amount of statistics that are collected for each player, team, game, and season. Data mining can be used by sports organizations in the form of statistical analysis, pattern discovery, as well as outcome prediction. Patterns in the data are often helpful in the forecast of future events. Similarly, researchers indicate that data mining can be used on physical aptitude test data in order to predict future physical performance. Data mining software is used to link test data with original data of particular event and helps to predict their actual performance in a required fitness class. Some of the work done in this field is presented below:

Berry & Linoff (2004), presented Sport organizations to acquire benefits by using Data Mining techniques which helps in deciding which product should have a promotion, the segmentation of customers, and the identification of the relationships among product items [10]. Robert P.etal (2006), applied data mining tools like WEKA and RAPID MINER in the field of sport. During this method users can run their data through one of the built-in algorithms, see what results come out, and then run it through a different algorithm to see if anything different stands out. As these programs' are available in the form of open source in nature, that's why the users are frequently to modify the source code, so that other can get the updated information [6]. Solieman, O.K(2006), used Data mining based Apriori and Decision Tree algorithm, for scouting, prediction of performance, selection of players, coaching and training and for the strategy planning [7]. Chodavarapu Y (2006), reviewed data mining techniques to determine the best or the most optimal squad to represent a team in a team sport in a season, tour or game [4].Smith, L.et al (2007), has presented data mining technique to predict Cy Young Award Winners. The award is given annually to the best pitcher in the major league of baseball. The award is based largely on statistics compiled over the course of the baseball season. A Bayesian classifier is developed to predict Cy Young Award winners in American major league baseball [5]. Lan Yu (2009), has presented an attempt by applying Data mining based association rule on test data of PHS using Microsoft Association Rules algorithm and SQL Server 2005. In the experiment, the grades of vital capacity, grip strength, standing long jump and step test of a student are used for input attributes, and total score of the student is used for prediction attribute. As a result, a lot of useful rules are generated. The grade of standing long jump is the most mportant influence factor on total score of a student [3]. dravko Ivankovic et al (2010), have presented an appliance of neural network in Basketball scouting. Data from the first B basketball league for men in Serbia, for seasons 2005/06,2006/07,2007/08,2008/09 and 2009/10. During these five seasons, a total of 890 games were played. These data were analysed using the feed-forward technique in neural network and using association rule on nonlinear data of basketball. As a result two-point shots under hoop and defensive rebound in basketball are considered as the most important elements of the game [2]. Kaur Parneet et al (2011), proposed an approach for the designing of DSS by performing the analysis of student Physical Fitness using Data Mining Algorithms i.e. Association Rule (Apriori Algorithm), Classification (Decision Tree) and Clustering. Along with that a comparison analysis is also presented between two tools WEKA and

TANAGRA by using a common test data file. In this analysis, various parameters i.e. Height, Weight, BMI, Speed Test, Balance Test, Flexibility Test, Medical Test of students is used for input attributes. The score of the student is used for prediction attribute. As the results, lots of useful rules are generated for prediction of physical fitness using these attributes which result in generating height and weight as the most important influence factor on total score of a student [1].

VI. ANALYSIS OF PHYSICAL HEALTH STANDARD USING ASSOCIATION RULE MINING ALGORITHM AND MATLAB FUZZY INFERENCE SYSTEM

In the previous work [1] data is applied using TANAGRA and WEKA but rules derived does not generate very useful results for future predictions. Due to drawbacks of earlier work, a new approach is required. In this paper, new fuzzy based Association Rule are derived using FIS. A problem of traditional association rule mining algorithms is that not every kind of data can be used for mining. In this analysis the PHS data of students from age 19 to 24 is applied on MATLAB Mamdani Fuzzy Inference System FIS [14] which helps in generating very good rules.

A. METHODOLOGY USED:

1. The synthetic data containing various parameters are processed first.
2. On the basis of the proposed model requirement fuzzy based items are derived from this processed data.
3. Then for each fuzzy based input membership functions are generated like for example three height membership functions are derived with different ranges i.e. for hght1 range is 155 metre to 160 metre, for hght2 range is 160 metre to 170 metre and for hght3 range is 170 metre to 175 metre.
4. Similarly for fuzzy based weight input, six weight membership functions are derived.
5. For output like BMI, Underweight, Normal and Overweight BMI fuzzy based items are formed.
6. Then for each membership functions formed during above steps, fuzzy rules are generated by using FIS Rule Editor. Some the rules formed are:

If Age (19-24) And Gender (Male) And Height (155-160) And Weight (51-60) Then BMI (Normal)

If Age (19-24) And Gender (Male) And Height (155-160) And Weight (40-51) Then BMI (Under-Weight)

If Age (19-24) And Gender (Male) And Height (155-160) And Weight (65-68) Then BMI (Over-Weight)

7. Results are derived on the basis of rules formed.
8. Results are viewed by using FIS Rule Viewer and Surface Viewer.
9. The same steps are followed for generation of membership functions for Endurance Test, Flexibility Test, Push-Up Test, Basketball, Hockey and Cricket game parameters.
10. Finally score based fuzzy rules are derived for each game, which helps in categorising players according to their physical fitness for particular game

VII. PROPOSED MODEL

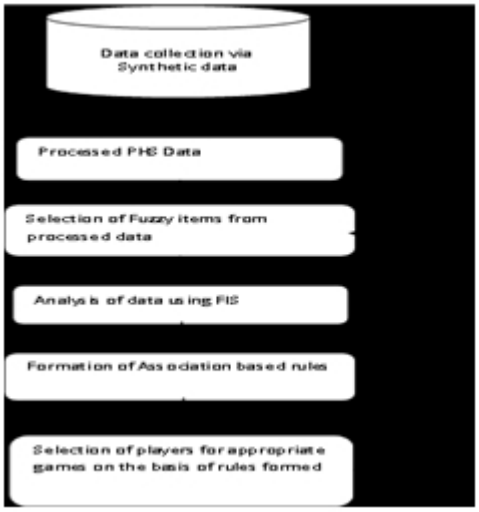

Figure1: Flow Chart of Analysis

VIII. RESULTS

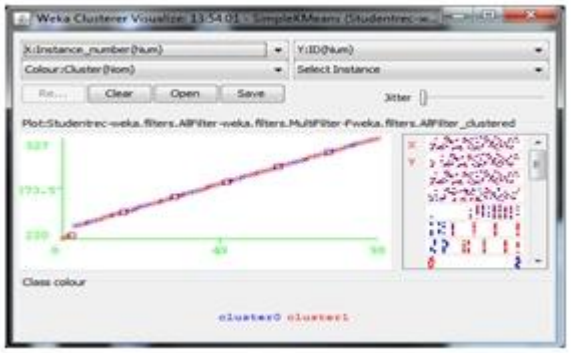

Figure 2: Association Rule generated through WEKA [1]

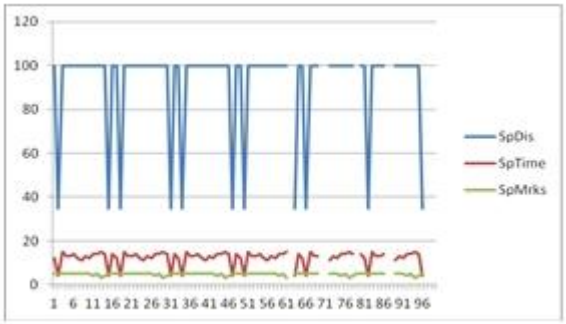

Figure 3: Graph formation for Endurance Test Parameters [1]

Figure 2 & 3 shows the results of synthetic data given in the appendix, applied on WEKA for generation of rules

by using Association based algorithm. Figure 2 shows the formation of frequent itemsets in the form of clusters.

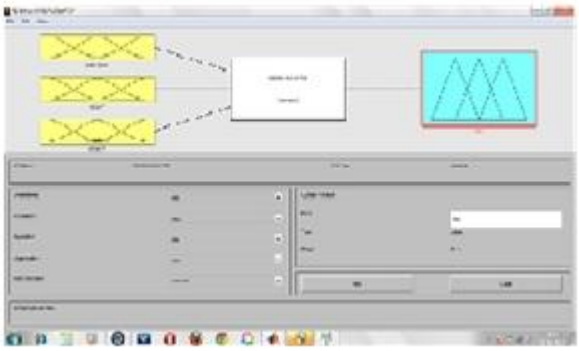

Figure 4: Membership functions for BMI in FIS Editor

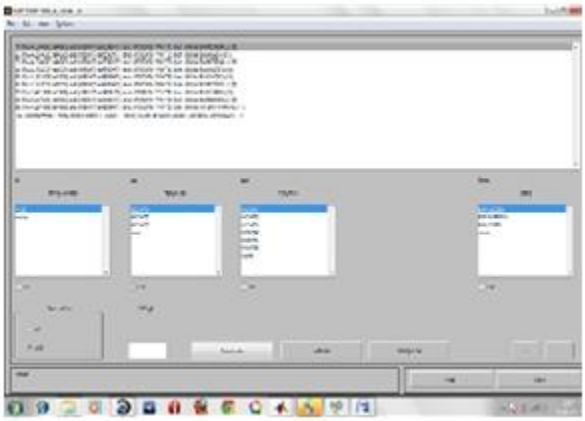

Figure 5: Rules formed for BMI using Rule Editor

Figure 4 shows the inputs i.e. weight and height and output as BMI applied on FIS. Figure 5 shows the rules generated from various inputs and outputs.

Figure 6: Graphical Rule View for BMI through Rule Viewer

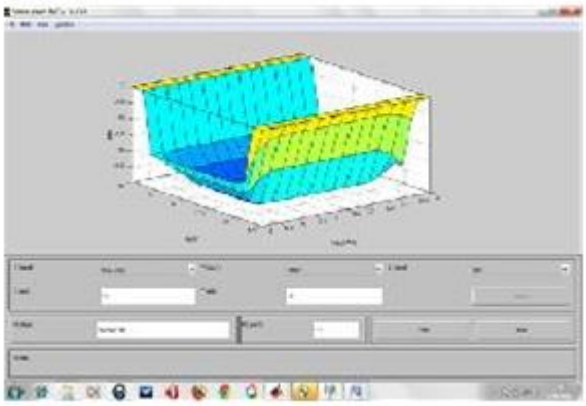

Figure 7: Surface View for BMI

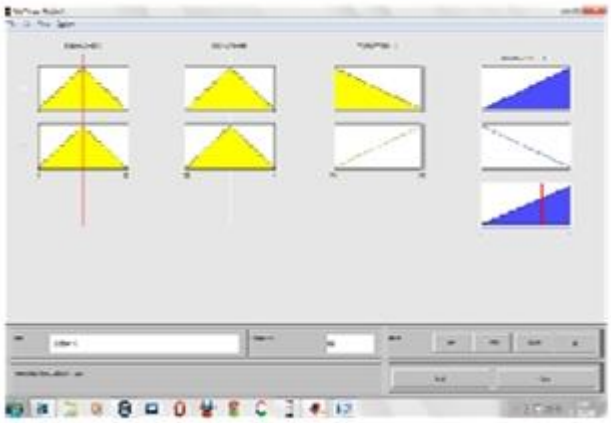

Figure 8: Graphical Rule View for Push-UP Test in range of 4-5

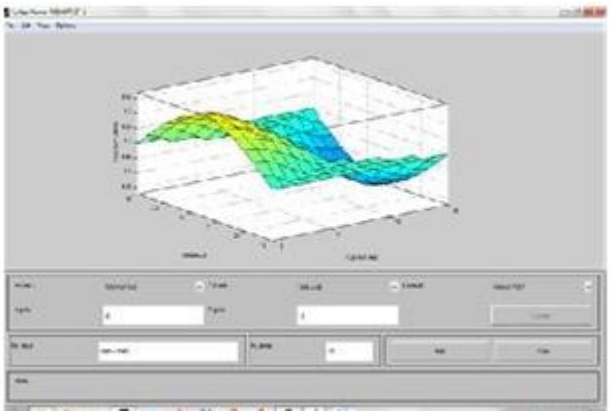

Figure 9: Surface View of Push-Up Test

The above Figure 8, Figure 9 shows the Push-Up Test Graphical view. Scale is formed in Figure 8 Rule View for no. of Push-ups in particular time for male candidate in the age group of 19-24.The final scale in blue show the Push-up Test Score which is in the range of 4-5 for eligible players. Figure 9 shows the 3 D Surface view of Push-Up Test Score.

Figure 10: Rule View of Basketball player selection

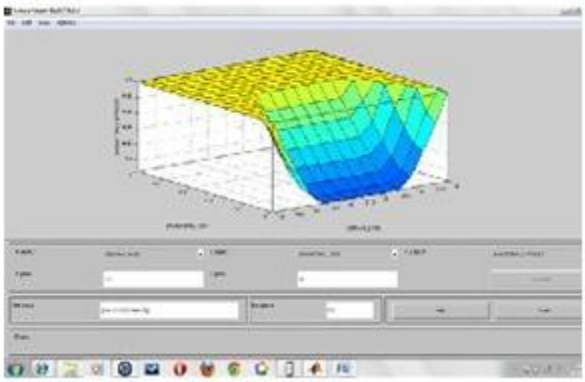

Figure 11: Surface View for Basketball player selection

The Figure 10 & Figure 11 shows again the Rule View and Surface view from the rules generated through Rule editor for the selection of those players which are suitable for playing Basketball. Similar results are derived for Endurance Test, Flexibility Test, Medical Test and other sports selection of players like Cricket, Hockey etc. The above results give very good and clear results which are absent in WEKA. These rule based results help the researchers in decision making and predictions for measurement of physical health standards and selection criteria's used in various sports.

VII. CONCLUSION AND FUTURE SCOPE

A fuzzy taxonomic structure is made for the items as some of the items are fuzzy like parameters of BMI calculation, Endurance test, Push-Up test etc. The rules derived from the above analysis are in the form of If-Then which are also formed through Association Rule algorithms. It is therefore necessary to provide a definition of association rules for the case of a database containing tables with fuzzy attributes. The work has not been analysed using other Data Mining algorithms which can be future scope of this paper. The discovery of fuzzy association rules for tables formed from data is considered an interesting and important research problem since apart from sports it can also be applied in various other fields also. Several research efforts showed that previously discovered association rules can benefit greatly to discover an algorithm that can produce new set of fuzzy association rules from tables.

ACKNOWLEDGEMENT

The proposed analytical model of PHS using Association Rule Algorithm with FIS is truly the result of many who belong to the list of author. The author like to express their thanks to Dr Sawtantar Singh and Dr Gurpreet Singh for their continuous encouragement guidance and support during this work.

REFERENCES

[1] Parneet Kaur, Bindu Rana (2011),"Analysis of Student Physical Health Data using Data Mining Algorithm", Proceeding of International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.6 No.18, pp.2325-2326.

[2] Zdravko Ivanković, Miloš Racković et al (2010), "Appliance of Neural Networks in Basketball Scouting", Proceedings of Acta Polytechnica Hungarica Journal, Vol. 7, No. 4, pp. 167-180.

[3] Lan Yu (2009), "Association Rules based Data Mining on Test Data of Physical Health Standard", Proceedings of International Joint Conference on Computational Sciences and Optimization, pp. 322-324.

[4] Chodavarapu Y. (2007), "Using data-mining for effective (optimal) sports squad elections".WebSite :http://insightory.com/view/74//using_datamining_for_effective_(optimal)_sports_squad_selections.

[5] Smith, L., Lipscomb, B., and Simkins, A. (2007), "Data Mining in Sports: Predicting Cy Young Award Winners". Proceedings of Journal of Computer Science, Vol. 22, pp. 115-121.

[6] Robert P. Schumaker ,Osama K. Solieman ,Hsinchun Chen,(2006).Contents of an article published in Springer.

[7] Solieman, O. K. (2006), "Data Mining in Sports: A Research Overview, A Technical Report, MIS Masters Project.

[8] Jadhav, S. R., and Kumbargoudar, P., (2007), "Multimedia Data Mining in Digital Libraries: Standards and Features". Contents of an article, pp. 54-59.

[9] Shitong W., Chung KFL., Hongbin S.(2005),"Fuzzy taxonomy, quantitative database and mining generalized association rules". Proceedings of Intelligent Data Analysis Jornal, Vol-9, issue-2, pp 207-217.

[10] Berry, M. J. A., & Linoff, G. S. (2004), "Data mining techniques: For marketing, sales, and customer relationship management", Book (2nd ed.). Indiana: Wiley Publishing, Inc.

[11] Saban, K. A. (2001),"The data mining process: At a

critical crossroads in development", Proceedings of Journal of Database Marketing, Vol.8 (2), pp.157-167.

[12] T.Anu Radha, P.Lavanya (2000)," Recent Trends in Parallel and Distributed Apriori algorithm", International Journal of Engineering Research and Applications, Vol. 1, Issue 4, pp.1820-1822, ISSN: 2248-9622.

[13] Savasere A, Omiecinski E, and avathe S. (1995), "An efficient algorithm for Mining Association Rules in large databases", Proceedings of 21st VLDB Conference.

[14] Agrawal, R. & Srikant, R. (1994),"Fast Algorithms for Mining Association rules". In Proc. 20th International Conference on Very Large Data Bases, pp. 487-499.

Tool Used:

[15] MATLAB Fuzzy Inference System FIS, Graphic User Interface.

Appendix: Synthetic Data of PHS

UserID	Stuname	Age	Height	Weight	Gender	MedicalTe	SpDis	SpTime
1001	Arjun	21	160	56	M	OK	100	12
1002	Anant	18	165	60	M	OK	35	4
1003	Avika	23	155	50	F	OK	100	15
1004	Anjali	21	160	55	F	OK	100	13
1005	Ananya	19	157	50	F	OK	100	13
1006	Akansha	23	160	51	F	OK	100	14
1007	Arpit	23	170	65	M	OK	100	12
1008	Ajay	22	175	65	M	OK	100	11
1009	Bindu	21	158	48	F	OK	100	13
1010	Bandini	21	155	48	F	OK	100	12
1011	Balbir	23	175	60	M	OK	100	14
1012	Chetna	17	150	47	F	OK	100	14
1013	Chirag	21	160	55	M	OK	100	15
1014	Anil	24	170	65	M	OK	100	14
1015	Amit	24	175	68	M	OK	35	4
1016	Arnav	20	160	50	M	OK	100	14
1017	Arpit	21	157	56	F	OK	100	12
1018	Ramesh	18	160	60	M	OK	35	4
1019	Nimrat	23	170	50	M	OK	100	15
1020	Pallivi	21	175	55	F	OK	100	13
1021	Payali	19	157	50	F	OK	100	13
1022	Payal	23	160	51	F	OK	100	14
1023	Ansh	23	165	65	M	OK	100	12
1024	Daksh	22	155	65	M	OK	100	11
1025	Divya	21	160	48	F	OK	100	13
1026	Pari	21	157	48	F	OK	100	12
1027	Virat	23	170	60	M	OK	100	14
1028	Vidya	17	155	47	F	OK	100	15
1029	Girish	21	175	55	M	OK	100	14
1030	Rajat	24	160	65	M	OK	100	4
1031	Nishant	24	170	68	M	OK	35	14
1032	Mayank	20	175	50	M	OK	100	12
1033	Krishan	22	160	56	M	OK	100	4
1034	Sandeep	18	165	60	M	OK	35	15
1035	Rakhi	22	155	50	F	OK	100	13
1036	Seema	23	160	55	F	OK	100	13
1037	Sashi	25	167	50	F	OK	100	14
1038	Radha	19	160	51	F	OK	100	12
1039	Gaurav	18	170	65	M	OK	100	11
1040	Nandish	25	175	65	M	OK	100	13
1041	Summi	21	158	48	F	OK	100	13

UserID	Stuname	BalTime	BalMrks	HockeyMr	BasketBall	Boxing	Cricket	BMI	BMIMarks	BMIScore
1001	Arjun	34	3	4	5	5	4	21.875	5	
1002	Anant	55	5	0	4	0	0	22.03	5	
1003	Avika	40	1	4	5	4	5	20.81	5	
1004	Anjali	35	3	5	4	4	5	21.48	5	
1005	Ananya	45	4	0	5	0	0	20.28	5	

1006	Akansha	40	1	4	4	0	5	19.92	5
1007	Arpit	35	3	5	0	4	0	22.49	5
1008	Ajay	40	1	5	5	0	0	21.22	5
1009	Bindu	40	1	5	0	0	5	19.22	5
1010	Bandini	45	4	5	4	5	3	19.97	5
1011	Balbir	35	3	5	4	0	0	20.76	5
1012	Chetna	30	3	5	0	5	0	19.56	5
1013	Chirag	25	3	0	5	0	4	21.48	5
1014	Anil	45	4	5	5	4	4	22.49	5
1015	Amit	35	3	4	0	0	3	22.2	5
1016	Arnav	30	3	3	0	0	0	20.28	5
1017	Arpit	34	3	4	4	3	4	21.875	5
1018	Ramesh	55	5	3	4	4	0	22.03	5
1019	Nimrat	40	1	3	5	2	5	20.81	5
1020	Pallivi	35	3	5	4	3	5	21.48	5
1021	Payali	45	4	3	5	3	0	20.28	5
1022	Payal	40	1	4	3	0	5	19.92	5
1023	Ansh	35	3	5	2	4	3	22.49	5
1024	Daksh	45	1	5	4	3	0	21.22	5
1025	Divya	40	1	5	3	2	5	19.22	5
1026	Pari	35	4	5	4	5	3	19.97	5
1027	Virat	40	3	5	4	3	0	20.76	5
1028	Vidya	35	3	5	2	4	0	19.56	5
1029	Girish	40	3	4	3	0	4	21.48	5
1030	Rajat	40	4	4	3	4	4	22.49	5
1031	Nishant	45	3	4	3	4	4	22.2	5
1032	Mayank	35	3	3	5	0	3	20.28	5
1033	Krishan	30	3	4	0	3	5	21.875	5
1034	Sandeep	25	5	4	4	0	4	22.03	5
1035	Rakhi	45	1	4	0	4	5	22.81	5
1036	Seema	35	3	5	3	4	5	21.87	5
1037	Sashi	30	4	3	5	0	2	20.28	5
1038	Radha	25	1	4	0	0	4	19.92	5
1039	Gaurav	45	3	5	0	0	5	22.49	5
1040	Nandish	35	1	5	4	0	0	21.22	5
1041	Summi	30	1	4	0	0	4	19.22	5