RESEARCH INVENTY: International Journal of Engineering and Science
ISSN: 2278-4721, Vol. 1, Issue 2 (Sept 2012), PP 35-41
www.researchinventy.com

Data Mining System Architecture for Training Plan Selection for the Swimmers

[1]Ms. D. P. Vaidya, [2]Dr. Sanjay U. Makh
[1](Assistant Professor, P. G. Department of Computer Science & Technology, Degree College of Physical Education, Shree Hanuman Vyayam Prasarak Mandal, Amravati (M.S.), India)
[2](Head of Department, Department of Electronics Technology, Takshashila Mahavidyalaya, Amravati (M.S.), India)

Abstract *– In Sports, vast amount and variety of data is generated at different stages. Sports are using this data to derive valuable information to improve performance, avoid injuries, organizations, researchers, coaches, players game planning, decision making and even for forecasting purpose. The Sports training/coaching required to be systematic and scientifically formulated approach to excel in sports. The aim of systematic Sports training is to produce best performance from a player. Performance of a player in competitive sports is joint effect of physiological, medical, psychological parameters, motor fitness, training/ coaching etc. A common training plan used during training may suite to some player in the group and improve their performance whereas it decreases the performance of some other group of players in the same team. Researches in the sport and physical education have proved that, intensity and type of training required to different players is different. Therefore a personalized approach in training/coaching will affect the performance in sports. A computer based coach assistant system is suggested and designed in this research.*

Keywords *– System architecture, Data mining, Sports Training, Time Series Analysis, Clustering.*

I. INTRODUCTION

Traditionally sports knowledge has been believed to be contained in the minds of its experts – the scouts, coaches, and managers. Sports organizations have now begun to realize that there is a wealth of knowledge contained in their data. The coaches who are in-charge of the team on the playing surface, as well as the general managers, who are in-charge of drafting or signing players, try to retrieve meaning and insight from the wealth of data for the scouts to evaluate future prospects and talent. Most in-house statisticians and analysts are helping the sports organization to gain valuable information from the data available in sports domain.

Hidden knowledge in the data gathered in Sports activity is required to be understood by the coaches and trainers, to apply it correctly to the training process of a particular sport. The Sports training is based upon many factors like: efficiency, endurance, skills, body types, socio-psychological parameters, nutrition, etc. The training should be individualized. This individualization helps to achieve maximum performance from each individual player. Swimming is a competitive sport in which players' performance mainly depends upon the physical fitness, skill and training of the player. The skill and training plays vital role in performance and personalization is required in this aspect. To ensure personalization it requires individual monitoring and evaluation which is quite impossible without use of any tool.

A data mining based system architecture is designed and the system is implemented which recommends and evaluates personalize training plan for individual player to enhance his/her skill and performance. This tool guarantees the improvement in the performance of the swimmers as the decisions of assigning training plan is based on scientific principles.

II. Review Of Previous Work

The decisions concerning the nature of sports training made with each individual athlete in mind to maximize performance. A coach must always consider that each athlete be treated independently as they are quite different from each other. They have different performance and fitness attributes, life-styles and nutritional preferences, and they respond to the physical and social environments of training in their own unique ways. Attempts to copy the training plan of champions, which is still a common practice among many coaches, will result in incorrect work load of training for most individuals. [1]

Relationship exists among Anthropometric Characteristics, Stroke Frequency and Stroke Length in Swimmers. [2] Several authors have been investigating the relationship among speed (V), stroke length (SL) and stroke frequency (SF) in competitive swimming for different distances with different swimming styles. The

researcher identified which of the anthropometric measures are more related to stroke frequency (SF100) and stroke length (SL100) in the 100 meter freestyle events for Brazilian elite swimmers. [2]

The following conclusions have been drawn from the results of this investigation:

i. A negative correlation prevailed for the variable SF.
ii. A positive correlation prevailed for the variable SL.
iii. The athletes who use a greater stroke frequency in the 100 meter freestyle tend to be shorter, with shorter arms and legs, narrower thorax and deeper abdomen, thicker arms and legs and larger feet.
iv. The athletes who use a larger stroke length in the 100 meter freestyle tend to be taller, with longer arms, narrower pelvis and wider thorax, thinner arms and legs and smaller feet. [2]

Anthropometric Measurements is significant characteristic in choosing players in many sport activities. [3] Anthropometric measurements affect effectively in the level of athletic achievement during the competitions, as there is a relationship between measurements and physical multi-level performance skills. It was suggested that the coach and those preparing for a software selection and choice of players must consider following points:

- Find measurement of lengths and circumferences, as well as the amount of obesity of the body.
- Conduct research on the strength of relations between these measurements and the relative weight of impressionist in the level of skill performance of the player. [3]

There are various training plans available. These training plans are assigned to the swimmers with different somatotypes and their performance can be tested. Sample Training plans are given below:

First Week - Day 1	
Warm up	400 m slow and easy, working on your water feel. Alternating Freestyle, Breaststroke and Backstroke.
Drills	Freestyle drills for 400 m, pick them out among those explained on the freestyle swimming drills post.
Main Set	2 X 400 m freestyle, medium pace with flip turns and dolphin kicking
Exercises	100 m breaststroke only leg with kickboard , 200 m freestyle only leg with kickboard , 100 m butterfly only leg with kickboard
Cool Down	200 m slow swimming

Table. 1: Day wise Training Plan for Swimmer [4]

Set	Action	Training Category
Warm up	400 free 200 back 200 back/free (alternating 5 strokes back, 5 strokes free with Han's paddles)	Rec i.e. Pulse rate upto 120, rest as per choice, velocity of 80% threshold speed
Kicking	3 x 20 (short fins) on 3 min	EN1 i.e. Pulse rate 120-140, 10-30 seconds rest, velocity of 85% threshold speed
Cross-pool explosive burst	8(4 x 15 or pool width) on 15 s. kick on odd number underwater back butterfly in torpedo position. Swim even numbers in sprint butterfly. Rest 30 sec. between sets	SP3 i.e. Pulse rate max, 1:2 rest ratio, velocity of 100% to 110% Max velocity
Stroke Technique	3 x 100 with the Gripper on 1.5 min. Count strokes and hold the number of strokes for both 50 s even on each swim.	EN1 i.e. Pulse rate 120-140, 10-30 seconds rest, velocity of 85% threshold speed
Build set	3 x 100 on 1.5 min Descend 1-3 to a heart rate count of 170	
Main set	25 x 100 on a send-off that permitsabout 10 s rest. Maintain a heart rate of 170.	EN2 i.e. Pulse rate 130-170, 10-30 seconds rest, velocity of threshold endurance speed
Recovery	Easy 200	Rec i.e. Pulse rate upto 120, rest as per choice, velocity of 80% threshold speed
Sculling/Pull set	12 x 100 on 1:40 min. (4 with the Gripper, 4 with Han's paddles, 4 regular). Pull 75, scull 25 on the odd numbers; scull 25, pull 75 on even numbers	EN1 i.e. Pulse rate 120-140, 10-30 seconds rest, velocity of 85% threshold speed

Table 2: Components of Sample Practice session [5]

Sports is known for the vast amounts of statistics that are collected for each player, team, game, and season. This results in information overload for those trying to derive meaning from the statistics. Hence, sports are ideal for designing data mining tools and techniques. [6] 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. A pilot program was started in 2002, which used software to help predict player injuries by collecting data from workouts over a period of time. The biomedical tool created by Computer Associates produces predictions from the medical statistics amassed for each player. Since athletes are their biggest investments, teams are hoping that prediction of injury will help save millions of dollars. [6]

Clustering techniques is used for modeling performance and to analyze sport physiological data collected during incremental tests to support the planning of training sessions, to provide a tool for athlete self-evaluation. [7] Modeling athlete performance to analyze the progress of a test session, automatically assign the tested athlete to a group of athletes which are similar to him/her with respect to physical parameters and development of the test, and evaluate these groups with respect to two quality indexes of the performance of the athlete, whose real value is known only at the end of the test. It provides a continuous characterization of the progress of the test. [7]

Almost all managerial decisions are based on forecasts. [8] Every decision becomes operational at some point in the future, so it should be based on forecasts of future conditions. Forecasts are needed throughout an organization and they should certainly not be produced by an isolated group of forecasters. Neither is forecasting ever "finished". Forecasts are needed continually, and as time moves on, the impact of the forecasts on actual performance is measured; original forecasts are updated; and decisions are modified, and so on. [8] A time series is a set of ordered observations on a quantitative characteristic of a phenomenon at equally spaced time points. One of the main goals of time series analysis is to forecast future values of the series. A trend is a regular, slowly evolving change in the series level. In Time-Series Models, it is assumed that there is no information about the causality that affects the variable we are trying to forecast. Instead, the past behavior of a time series is examined in order to infer something about its future behavior. Forecasting is a prediction of what will occur in the future, and it is an uncertain process. Because of the uncertainty, the accuracy of a forecast is as important as the outcome predicted by the forecast. [8]

III. THE SYSTEM ARCHITECTURE

The architectural diagram for the system is given below:

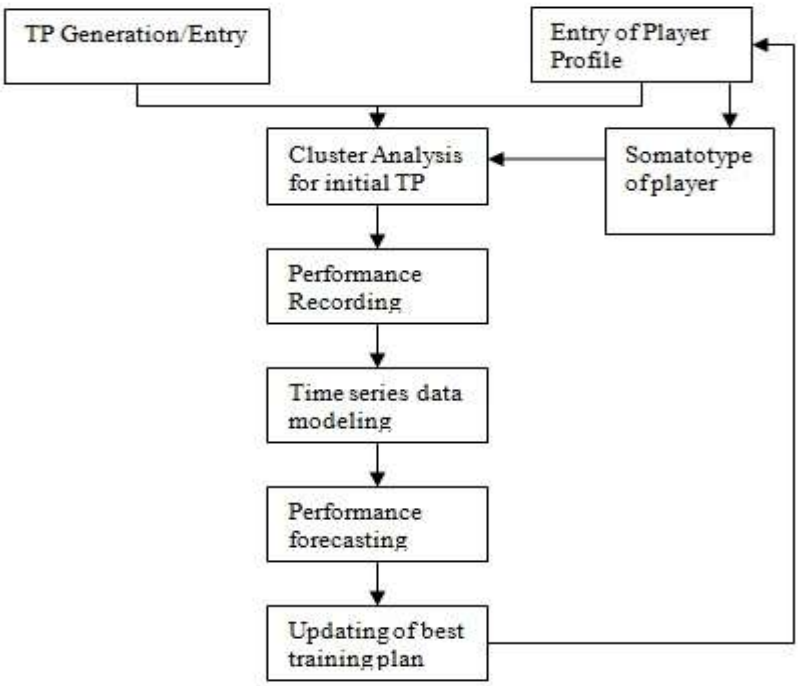


Fig. 1: Architectural Diagram for the system

The coaches give training to the players according to some fixed training plan; the plans are exhaustive and aimed to improve the performance of the swimmer in the competitive sports. The exhaustive training plans may or may not suit to all the swimmers as they have different body type and anthropometric measurements and physical fitness. Based upon the previous performance data against training plan, cluster of swimmers is formed. The Nearest Neighbor Algorithm is applied for forming the cluster. The training plans which were best fitted to the swimmers in the cluster are found and suggested to the new swimmer. Time series analysis is performed on the data gathered for each swimmer for assigned training plan, the future performance of the swimmer is then predicted against training plan. The trend in the performance is used for taking decision like whether to continue or discontinue the training plan assigned to the swimmer. This saves time of experimentation of training plans on new swimmer as we have huge amount of practice session data available in the tool.

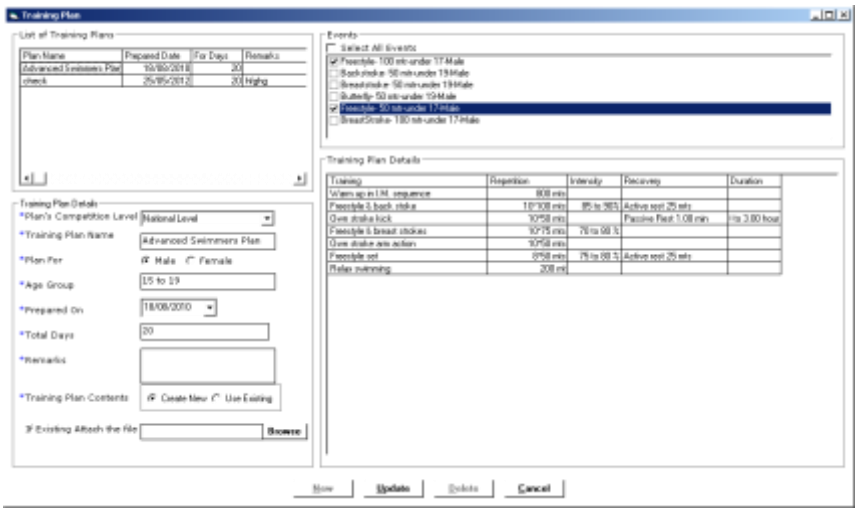


Fig. 2: Form for creating Training Plan Format.

If the format for training plan does not suffice any of the coach's need, coach can input their own created file from MS Word, MS Excel, PDF, etc. through the provision given in the screen to attach existing files.

The swimmer's personal information, the practicing events along with various anthropometry measurements are recorded.

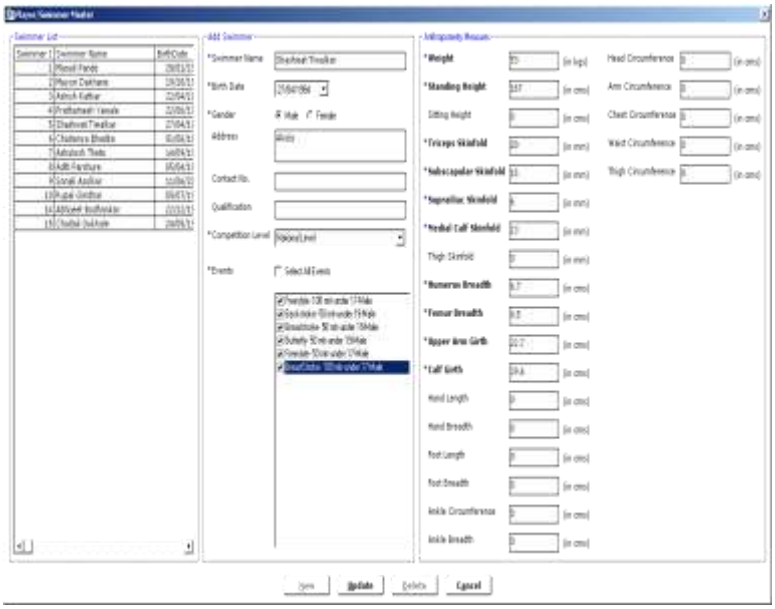
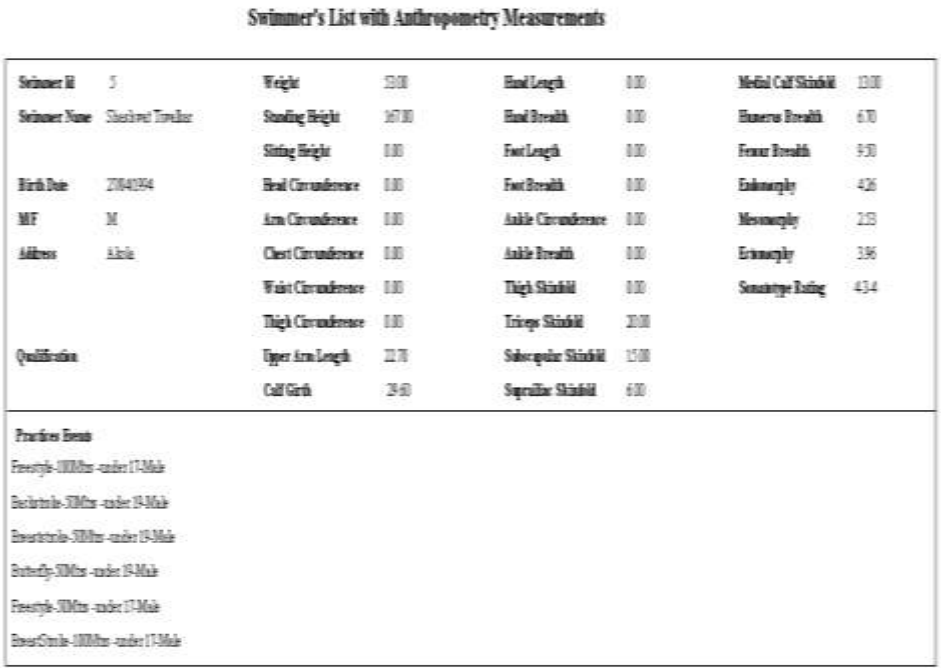


Fig. 3: Form for Swimmer's personal profile, anthropometric data and event he/she practices.

Based upon above calculation method, the system generates following type of somatotype output:



Swimmer's List with Anthropometry Measurements

Swimmer Id	5	Weight	53.00	Hand Length	0.00	Medial Calf Skinfold	13.00
Swimmer Name	Shashwat Tiwalkar	Standing Height	167.00	Hand Breadth	0.00	Humerus Breadth	6.70
		Sitting Height	0.00	Foot Length	0.00	Femur Breadth	9.30
Birth Date	27/04/1994	Head Circumference	0.00	Foot Breadth	0.00	Endomorphy	4.26
M/F	M	Arm Circumference	0.00	Ankle Circumference	0.00	Mesomorphy	2.53
Address	Akola	Chest Circumference	0.00	Ankle Breadth	0.00	Ectomorphy	3.96
		Waist Circumference	0.00	Thigh Skinfold	0.00	Somatotype Rating	4.3-4
		Thigh Circumference	0.00	Triceps Skinfold	20.00		
Qualification		Upper Arm Length	22.70	Subscapular Skinfold	15.00		
		Calf Girth	29.60	Suprailiac Skinfold	6.00		

Practices Events

Freestyle-100Mtrs -under 17-Male

Backstroke-100Mtrs -under 19-Male

Breaststroke-100Mtrs -under 19-Male

Butterfly-100Mtrs -under 19-Male

Freestyle-100Mtrs -under 17-Male

Breaststroke-100Mtrs -under 17-Male

Fig. 4: Report showing Swimmer's information along with Anthropometry measurements and somatotype rating found. It also shows Events practiced by the swimmer.

Clustering determines the similarity among the data on predefined attributes. The data found to be most similar is grouped into clusters. For implementing clustering technique the Nearest Neighbor Algorithm is used. System intimates coach regarding best suited training plan for the player on the basis of existing cases in the system database. The system output is as give below:

Systems Suggestion for assigining Training Plan on Performance Evaluation of previous Swimmer's

Details of Swimmer under consideration

Swimmer Id	5	Weight	53.00	Hand Length	0.00	Medial Calf Skinfold	13.00
Swimmer Name	Shashwat Tiwalkar	Standing Height	167.00	Hand Breadth	0.00	Humerus Breadth	6.70
		Sitting Height	0.00	Foot Length	0.00	Femur Breadth	9.30
Birth Date	27/04/1994	Head Circumference	0.00	Foot Breadth	0.00	Endomorphy	4.26
M/F	M	Arm Circumference	0.00	Ankle Circumference	0.00	Mesomorphy	2.53
Address	Akola	Chest Circumference	0.00	Ankle Breadth	0.00	Ectomorphy	3.96
		Waist Circumference	0.00	Thigh Skinfold	0.00	Somatotype Rating	4.3-4
		Thigh Circumference	0.00	Triceps Skinfold	20.00		
Qualification		Upper Arm Length	22.70	Subscapular Skinfold	15.00		
		Calf Girth	29.60	Suprailiac Skinfold	6.00		

List of Swimmers matching the selection criteria :

Swimmer Id	14	Weight	61.00	Hand Length	0.00	Medial Calf Skinfold	27.00
Swimmer Name	Abhijeet Bodhankar	Standing Height	173.00	Hand Breadth	0.00	Humerus Breadth	6.20
		Sitting Height	0.00	Foot Length	0.00	Femur Breadth	8.90
Birth Date	22/12/1997	Head Circumference	0.00	Foot Breadth	0.00	Endomorphy	4.21
M/F	M	Arm Circumference	0.00	Ankle Circumference	0.00	Mesomorphy	2.01
Address		Chest Circumference	0.00	Ankle Breadth	0.00	Ectomorphy	3.59
		Waist Circumference	0.00	Thigh Skinfold	0.00	Somatotype Rating	4-2-4
		Thigh Circumference	0.00	Triceps Skinfold	15.00	Best Training Plan	Advanced Swimmers Plan
		Upper Arm Length	24.40	Subscapular Skinfold	18.00		
Qualification		Calf Girth	35.00	Suprailiac Skinfold	9.00		

System Suggest following Training Plan based on previous Performance Evaluation

Training Plan (s)

Advanced Swimmers Plan

Fig. 5: Suggestion by the system to assign Training plan to new swimmer

3.1 Performance recording

The performance of the swimmer is to be judged against the event and assigned training plan. Screen is provided to record the performance details of a practice session.

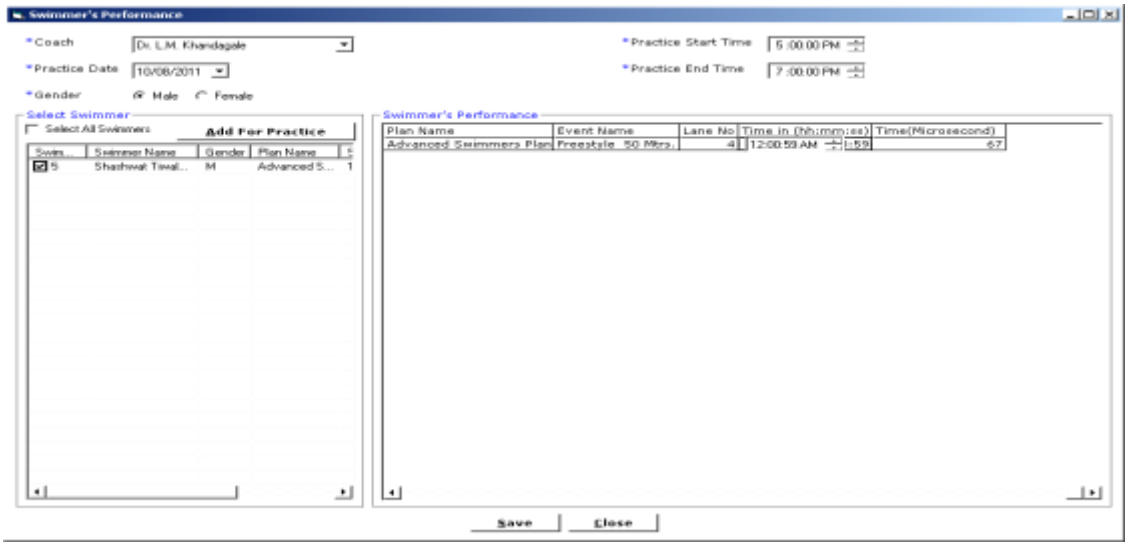


Fig. 6: Form for recording the performance of the swimmers

3.2 Forecasting of Player Performance

The graphical presentation of the output, well explain the trend of data values. A straight line is fitted for the performance and extended for the future values of the performance variable. The graph is as shown below:

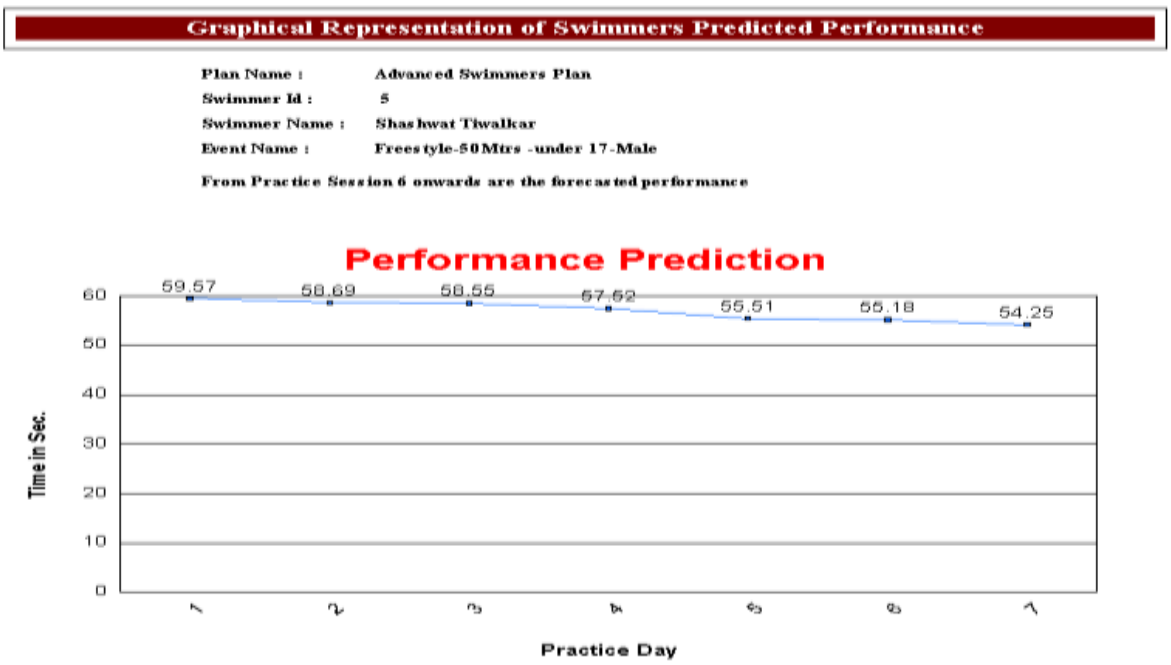


Fig. 7: Graphical representation showing Performance prediction.

After recording few completion timings of a swimmer against an assigned training plan, if it is found that the performance is not enhancing then the training plan can be discontinued. Another training plan can be assigned and again swimmer's performances timings can be recorded for the newly assigned training plan. For this task an input screen is provided.

IV. Conclusion

Various Sports activities are practiced around the world since ancient time. From the critical study it is observed that, body type, body posture along with proper training plays a vital role in the performance of the player in any game. The player with specific body type can excel in the specific sports/game. The scientific approach in systematic training process contributes lot in performance enhancement. Swimming is one of the oldest sport activities. In scientific training coaches gather various type of information about the player and monitor performance day-to-day during the training camps and thus variety of data available in swimming sports domain is increasing gradually. Sports organization, sports scientist, coaches and swimmers are now attracted to extract valuable information, hidden in this data for improvement of performance. The systematic sports training sessions, enhances the competitive sport performance in swimming. Coaches prepare training plans and implement them on the swimmers. Many times same training plan is assigned to all the swimmers in a specific age group. But in this research it is found that the similar training plan does not suit to all swimmers in specific age group as the swimmers are not having similar anthropometric, physiological and other related parameters, which contribute in sports/games performance. Thus it is required that training plans should be personalized.

Data mining technology is found useful in finding hidden, previously unknown and valuable information from large amount of data. Therefore, system is developed using this technology to help the coaches in scientific way to decide which training plan is suitable to swimmer and guarantee better performance. System performs performance forecasting based on the day-to-day training performance data. System evaluates cluster analysis to identify the cluster of players to which new player can be classified. When a new swimmer is recorded, a cluster of swimmers is formed based upon their anthropometric measurements or body type or somatotype information. The training plan which is found to be best suited to the swimmers in the cluster is then recommended this new player. Nearest Neighbor algorithm of Clustering technique is found to be useful for clustering of swimmers. To forecast the performance against each training plan, time series analysis is performed. This system is found useful in guiding the coaches in the training process and ultimately helps to improve the swimmer's performance. Using this system coaches can select a training plan, which suit each individual player according to his/her anthropometric measures.

References

[1]. B. S. Rushall and F. S. Pyke, (1990), The Principle of Individuality. [online] Available from: http://coachsci.sdsu.edu/csa/vol34/rushall3.htm (Accessed 12 May 2012)

[2]. Silvio Soares dos Santos, Hartmut Riehle, (1998), Relationship among Anthropometric Characteristics, Stroke Frequency and Stroke Length in Brazilian Elite Swimmers. [online] *16 International Symposium on Biomechanics in Sports*. Available from: http://w4.ub.uni-konstanz.de/cpa/article/view/988 (Accessed 8 April 2012)

[3]. Ahmed Nabieh Ibrahim Mohamed**,** Anthropometric Measurements as a Significant for Choosing Juniors in Both Volleyball and Handball Sports (Factorial Analysis Study). [online] *World Journal of Sport Sciences*. Available from: http://idosi.org/wjss/3(4)10/7.pdf 988 (Accessed 17 April 2012)

[4]. Gabriel, The "4 weeks swimming training plan". [online] Available from: http://effectivewaterfitness.com/wp-content/uploads/2009/03/the_4_weeks_swimming_training_plan_v1. pdf (Accessed 17 May 2012)

[5]. D. Hannula, (1995), *Coaching Swimming successfully* (USA, Human Kinetics, 1995)

[6]. O. K. Solieman, Data Mining in Sports: A Research Overview. [online], Available from: http://ai.arizona.edu/mis480/syllabus/6_Osama-DM_in_Sports.pdf (Accessed 8 April 2012)

[7]. Y. LI, S. Chiusano and Ing. Vincenzo D'Elia., Modeling Athlete Performance Using Clustering Techniques [online], Available from: http://www.academypublisher.com/proc/isecs10w/papers/isecs10wp169.pdf (Accessed 17 April 2012)

[8]. Professor H. Arsham, Time-Critical Decision Making for Business Administration. [online], Available from: http://home.ubalt.edu/ntsbarsh/stat-data/forecast.htm (Accessed 18 April 2012)

[9]. Lind, Marchal, Mason, Gupta, Kabadi, and Singh, Statistical Techniques in Business & Economics. Time Series and Forecasting, [online], Available from: http://highered.mcgraw-hill.com/sites/dl/free/0070951640/354829/lind51640_ch16.pdf (Accessed 3 May 2012)

[10]. Dr. S. P. Gupta, *Statistical Methods* (New Delhi, India, S. Chand & Sons Educational Publishers, 2007).

[11]. M. Dunhum, *Data Mining: Introductory and Advanced topics* (New Delhi, India, Dorling Kindersley (India) Pvt. Ltd, 2006)

[12]. A. K. Pujari, *Data Mining Techniques* (Hyderabad, India, University Press (India) Pvt. Ltd., 2001).