IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 3, Issue 1, January 2015 ISSN 2321-5992

A SURVEY AND ANALYSIS OF STUDENT INVOLVEMENT IN SPORTS

[1]S. SANDHIYA, [2] S. BABU, M.Phil., (Ph.D)

[1]M.Phil. (CS) Research Scholar SCSVMV University
Enathur, Kanchipuram

[2]Asst. Professor Department of CSA, SCSVMV University
Enathur, Kanchipuram

ABSTRACT

The present student generation are interested only in education that is, that gives good earnings. The students are totally not interested in other extra curricular activities particularly they are not interested in games and sports. Generally compare to boys, the girls are not much interested in games and sports. If at all some girls are interested in sports, their parents are hesitate to allow them to participate in the event. In the world, millions of peoples are affected by the diabetics and hypertension, the reason is lack of exercises. So this research has been conducted about the involvements of students in sports. Both boys and girls were brought in this research. The research study was conducted among the students of 9^{th} to 12^{th} std in the schools of rural and urban areas. Research was conducted with the structured questionnaire to collect the required primary data from the students, collected data were analysed. The present study conducted among 600 students in and around kanchipuram within tamilnadu to identify the student's involvements in sports and games.

1. INTRODUCTION TO STUDENT'S INVOLVEMENT IN SPORTS

Swami Vivekananda said "those who have sound body would have sound mind", so for good health and sound mind the sports and games are very important. In the world, millions of peoples are affected by the diabetics and hypertension, the reason is lack of exercises. Sports awareness is being reduced among the students and the parents. Now a days the parents are only focused on their children's education only. And they are continually taken care of their student's education. And most of the times they spent for their education and no one will care about their student's extracurricular activities. So this research is shown that the student's desire about the sports and student's favourite sports. And also this research classifies the students as male and female, and urban and rural area students. So we can understand the student's desire and the expectations about the sports among these various categories. This research is talks about the male and female student's characteristics in sports and also it differentiate the urban and rural area student's desire like their favourite sports and characteristics like their expectations about the sports.

1.1 OBJECTIVE OF THE STUDY

The Objectives that stimulate the study are set as follows:

- Assess the existing literature on student's involvement in sports that to predict the student's involvement.
- To analyse what are the factors to determine student's involvement in sports.
- To analyse the factors that the student's expectations about the sports and other activities whether they are interested or not.
- To suggest the factors to improve the student's physical training and activities.

1.2 STATEMENT OF THE PROBLEM

This research is talks about the male and female student's characteristics in sports and also it differentiate the urban and rural area student's desire, like their favourite sports and characteristics like their expectations about the sports. Through this research we can find out that the drawbacks of implementing the practice of sports and other activities to the students. Though the students are interested, the parents are not allowed them to play the games and to participate in games activities. And we can differentiate this problem by various categories.

2. REVIEW OF LITERATURE

[1] Bruce Kelley and Carl Carchia | ESPN The Magazine, The hidden demographics of youth sports, ESPN's summer 2013 Kids in Sports focus, we mined the often hidden-away data to paint as comprehensible a portrait of the nation's competitive youth sports landscape as we could. The Sports and Fitness Industry Association (SFIA), which employs tens of thousands of online interviews, tallies how many kids between 6 and 17 are regular/frequent (or what it calls "core") players of different sports. These "core" numbers make a decent stand-in for kids who play on organized teams, even though that's not the question asked; what's asked is whether a kid played that sport a minimum of 13 times a year in a sport like ice hockey or 26 times a year in a sport like soccer. SFIA gave ESPN The Mag custom data totaling up its 2011 participation.

[2] Matthias Chan Final Paper Independent Study A survey of sports in operational research, This article talks about the last 50 years of OR in sport and gives examples for the different veins that have been pursued in the last 50 years, as well as a prediction of what may be pursued in the near future. This paper first gives a working definition of a sport: A sport is a competitive but essentially non-hostile activity that necessarily involves some kind of physical activity. There are 4 main areas of study for OR in sport: analysis of tactics and strategy, scheduling, forecasting, and other miscellaneous papers.

[3] Robert P. Schumaker, Osama K. Solieman and Hsinchun Chen Sports Knowledge Management and Data Mining Vast amounts of sports data are routinely collected about players, coaching decisions and game events. Making sense of this data is important to those seeking an edge. By transforming this data into actionable knowledge, scouts, managers and coaches can have a better idea of what to expect from opponents and be able to use a player draft more effectively. With millions of dollars riding on the many decisions made within a sports franchise (Lewis, 2003), the sports environment is ideal for data mining and knowledge management approaches. While the application these approaches to the sports environment may be unique and the focus of this chapter, the topics of data mining and knowledge management should certainly be well known to the reader and form the basis of the approaches

[4] S Schneider, B seither, S tonges Sports injuries: population based representative data on incidence, diagnosis, sequelae, and high risk groups To generate national representative data on the incidence, diagnosis, severity, and nature of medically treated sports injuries and to identify high risk groups. The first national health survey for the Federal Republic of Germany, conducted in the format of a standardised, written, cross sectional survey in the period October 1997 to March 1999, gathered data on the incidence of accident and injury and information on social demographics, injury related disability/time off work, and injury location/setting. The net sample comprised 7124 people aged 18–79.

[5] Pamela Wicker & Christoph Breuer (2005) Usefulness of Data Mining in sport management research Sport clubs are often objects of investigation in sport management research. Usually descriptive analyses are carried out regarding the presentation of the studies' results. Thereby, differentiations between sport clubs according to number of members or offered sports are undertaken. However, these features are not always the distinctive ones. As sport clubs are very heterogeneous, it is difficult to find out structures beside the mentioned ones and which features are really the distinctive ones. Therefore a more explorative procedure of data analysis should be used. Using Data Mining, several procedures and algorithms are applied simultaneously which lead to the detection of patterns and structures in mass data that would not have been identified using conventional statistical analyses. All the experiments reported in this dissertation were conducted using the WEKA TOOL.

3. MATERIALS AND METHODS

Data Mining is the exploration and analysis of large sets, in order to discover meaningful patterns and rules. The key idea is to find effective ways to combine computers power to process data with the human eye's ability to detect patterns. The techniques of data mining are designed for work best with large data sets. It is the set of activities used to find new, hidden or unexpected patterns in data or unusual patterns in data. The WEKA workbench is a collection of machine learning algorithms and data pre-processing tools. Developed at the University of Waikato in New Zealand, WEKA stands for Waikato Environment for Knowledge Analysis. WEKA is open source software issued under the GNU General Public License. WEKA is a collection of machine learning algorithms for data mining tasks. It is used for research, education, and applications. Main features of WEKA are:

- Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods.
- Graphical user interfaces (incl. data visualization).

- Environment for comparing learning algorithms.

3.1 DATA PRE-PROCESSING

As is common in data mining, before running tests on data instances, it was necessary to clean and prepare our data for use into the WEKA workbench. An important piece here was the need to convert string data into nominal data from the ARFF file. This was done based upon the requirements constraints of the algorithms used, as they do not accept string data for processing. In addition, it was important to look at relevance of the attributes to remove redundant, noisy, or irrelevant features. In our data, I only removed two attributes which were students register number and their name. Then I choose replace missing values to replace all missing values (choose→filters→ unsupervised→attribute→ ReplaceMissing Values) for attributes. Replacing missing values places the distribution towards the mean value of the most frequent values for an attribute, and prevents the loss of information which might potentially be useful for learning.

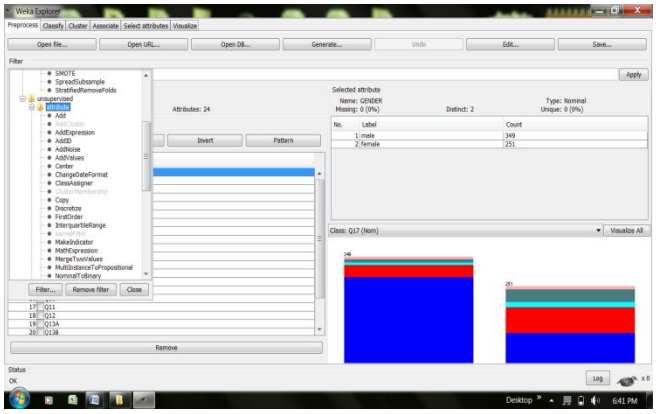


FIGURE 3.1: PRE-PROCESSING IN WEKA

3.2 CLUSTERING CONCEPT

Clustering is a division of data into groups of similar objects. Representing the data by fewer clusters necessarily loses certain fine details, but achieves simplification. It models data by its clusters. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept. From a practical perspective clustering plays an outstanding role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, CRM, marketing, medical diagnostics, computational biology, and many others. Clustering is widely used in gene expression data analysis. By grouping genes together based on the similarity between their gene expression proles, functionally related genes may be found. Such a grouping suggests the function of presently unknown genes.

3.3 Clustering Techniques

Traditionally clustering techniques are broadly divided in hierarchical and partitioning. Hierarchical clustering is further subdivided into agglomerative and divisive. The basics of hierarchical clustering include Lance-Williams formula, idea of conceptual clustering, now classic algorithms SLINK, COBWEB, as well as newer algorithms CURE and CHAMELEON. While hierarchical algorithms build clusters gradually (as crystals are grown),partitioning algorithms learn clusters directly. In doing so, they either try to discover clusters by iteratively relocating points between subsets, or try to identify clusters as areas highly populated with data. Partitioning Relocation Methods. They are further categorized into probabilistic clustering (EM framework, algorithms SNOB, AUTOCLASS, MCLUST), k-medoids methods (algorithms PAM, CLARA, CLARANS, and its extension), and k-means methods (different schemes, initialization, optimization, harmonic means, extensions).Such methods concentrate on how well points fit into their clusters and tend to build clusters of proper convex shapes. Partitioning algorithms of the second type are surveyed in the section Density-Based Partitioning. They try to discover dense connected components of data, which are flexible in terms of their shape. Density-based connectivity is used in the algorithms DBSCAN, OPTICS, DBCLASD, while the algorithm DENCLUE exploits space density functions. These algorithms are less sensitive to outliers and can discover clusters of irregular shapes. They usually work with low-dimensional data of numerical attributes, known as spatial data. Spatial objects could include not only points, but also extended objects (algorithm GDBSCAN).

K-mean Clustering

The conventional K-mean algorithm is based on decomposition, most popular technique in data mining field. The concept of K-Means algorithm uses K as a parameter, Divide n object into K clusters, to create relatively high similarity in the cluster and, relatively low similarity between clusters. And minimize the total distance between the values in each cluster to the cluster center. The cluster center of each cluster is the mean value of the cluster. The calculation of similarity is done by mean value of the cluster objects. The measurement of the similarity for the

algorithm selection is done by the reciprocal of Euclidean distance. That is to say, the closer the distance, the bigger the similarity of two objects, and vice versa.

Procedure of K-mean Algorithm

K-mean distributes all objects to K number of clusters at random; Calculate the mean value of each cluster, and use this mean value to represent the cluster; Re-distribute the objects to the closest cluster according to its distance to the cluster center; Update the mean value of the cluster, say, calculate the mean value of the objects in each cluster;

Calculate the criterion function E, until the criterion function converges. Usually, the K-mean algorithm criterion function adopts square error criterion, defined as: In which, E is total square error of all the objects in the data cluster, p is given data object, mi is mean value of cluster Ci (p and m are both multidimensional).

The function of this criterion is to make the generated cluster be as compacted and independent as possible [4].

Hierarchical Clustering

Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. The result of the algorithm is a tree of clusters, called dendrogram, which shows how the clusters are related. By cutting the dendrogram at a desired level, a clustering of the data items into disjoint groups is obtained [5].

Agglomerative (bottom up)

1. Start with 1 point (singleton).
2. Recursively add two or more appropriate clusters.
3. Stop when k number of clusters is achieved.

Divisive (top down)

1. Start with a big cluster.
2. Recursively divides into smaller clusters
3. Stop when k number of clusters is achieved.

General steps of Hierarchical Clustering

Given a set of N items to be clustered, and an N*N distance (or similarity) matrix, the basic
process of hierarchical clustering (defined by S.C. Johnson in 1967) is this [6]: Start by assigning each item to a cluster, so that if we have N items, we now have N clusters, each containing just one item. Let the distances (similarities) between the clusters the same as the distances (similarities) between the items they contain. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now we have one cluster less. Compute distances (similarities) between the new cluster and each of the old clusters. Repeat steps 2 and 3 until all items are clustered into K number of clusters. The merging criteria of clusters for hierarchical clustering are single link, average link and complete link use minimum, average and maximum distances between the members of two clusters, respectively.

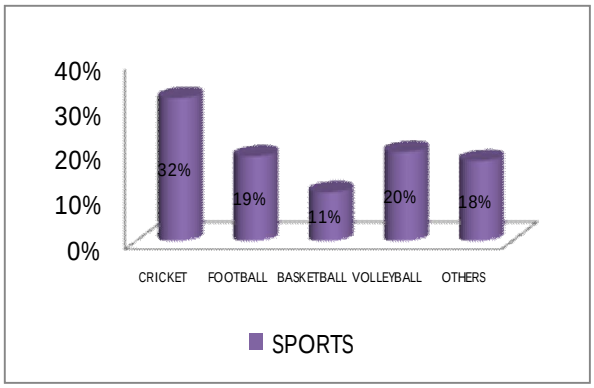
3.4 DATA COLLECTION

Data and sources of data: For this study a questionnaire was modelled and data regarding the opinion of the students in various schools in kanchipuram of tamil nadu. The research is fully based on the student's expectations about the sports and the student's favourite sports.

Sample selected for the study: The Questionnaire was circulated to 600 students from various schools. The sample respondents are selected on the basis of convenient sampling method.

4. BASIC ANALYSIS

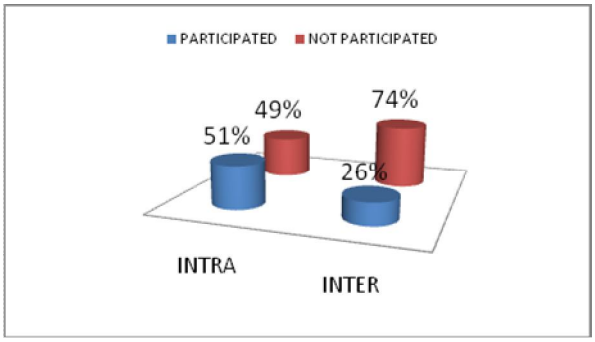
4.1 STUDENT'S PARTICIPATED SPORTS



The result of the analyse reveals that the students mostly interested in cricket next to cricket the students are interested in volleyball and football. 32% of students like cricket and 20% of students like football and volleyball. 18% of students interested in other sports like kho-kho, kabadi and etc.

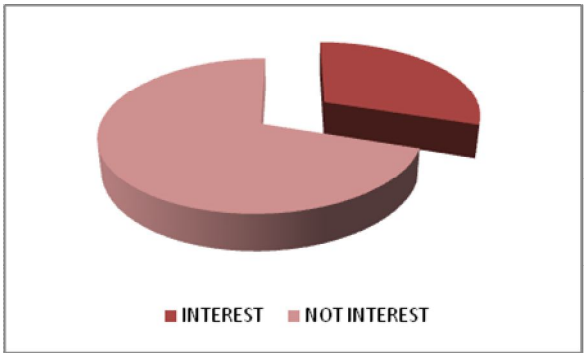
IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 3, Issue 1, January 2015 ISSN 2321-5992

4.2 INTRA AND INTER SCHOOL PARTICIPATION

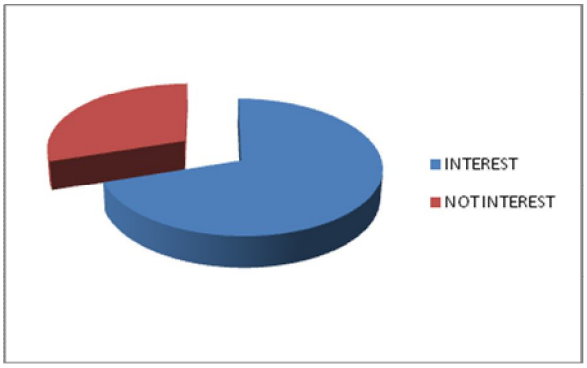


Most students wanted to participate in intra level competition compare to inter level competition. 51% of students are participated in intra level competitions and 26% of students are participated in inter level competition.

4.3 FEMALE STUDENT'S INVOLVEMENT IN SPORTS

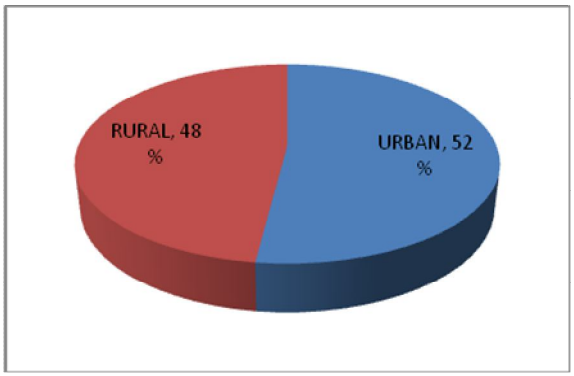


4.4 MALE STUDENT'S INVOLVEMENT IN SPORTS



The result of this analysis shows that comparing the male students the female students are less interested in sports activities. The physical education teachers are interested to develop the efficiency of the male students in sports compare to the female students. And most of the co-education schools physical teachers are male teachers only.

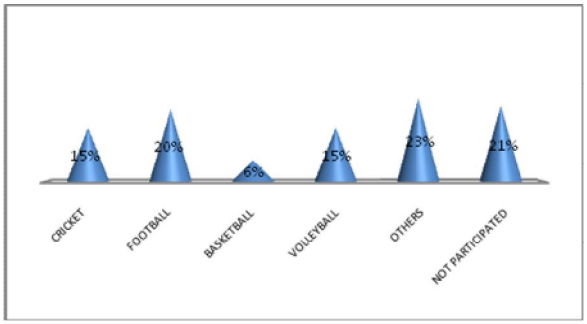
4.5 URBAN AND RURAL AREA STUDENT'S INTEREST IN SPORTS



As for as the interest of the student in sports concern 52% of urban area students interested and 48% of rural area students are interested in sports.

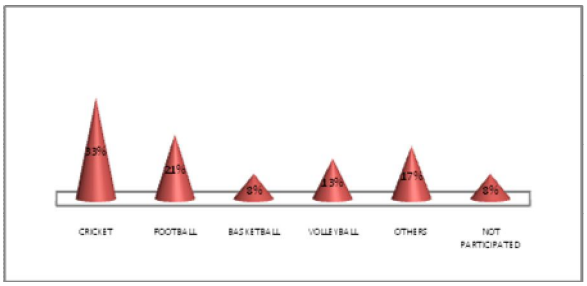
IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 3, Issue 1, January 2015 ISSN 2321-5992

4.6 COMPETITION WON BY URBAN STUDENTS:



Mostly urban area students won in the football matches next to that they won in cricket match. And 21% of students are not interested to participate in any competition. 23% of students won in other sports like kho-kho, kabadi and so on.

4.7 COMPETITION WON BY RURAL STUDENTS



Mostly the rural area students won cricket and football. 33% of rural area students won cricket and 21% of students won football match. 17% of students won in other sports like kho-kho and kabadi. And 8% of students don't like to participate in any sports competition.

5. USING WEKA ANALYSIS

The main aim to show the comparison of the different-different clustering algorithms on WEKA and find out which algorithm will be most suitable. All these algorithms are compared according to the following factors: size of dataset, number of clusters, and the classified attributes in clusters. Some conclusions that are extracted belong to the performance, quality, and accuracy of the clustering algorithms.

5.1 simplekmeans

Using simplekmeans algorithm, 600 instances and 24 attributes were used, the data was analysed to find out the student's involvements in sports based on their attributes. From this simplekmeans algorithm 43% of students among 600 are like to play sports. This analyse is done based on the gender and the area of the students. From the Simplekmeans algorithm, the category table shows that the each clusters have a superiority attributes. Using these results we can find the charecteristics of all clusters.

No. Of Iterations:7

Table 5.1 Simple kmeans category table

Attribute	Clus 0	Clus 1	Clus 2	Clus 3	Clus 4	Tot
Gender	X		X	X		3
Std						0
Home town		X	X	X	X	4
School		X		X		2
Income	X		X		X	3

5.2. Farthestfast algorithm

Using Farthestfast algorithm, 600 instances and 24 attributes were used, the data was analysed to find out the student's involvements in sports based on their attributes. From the FarthestFirst algorithm cluster 1 and 3 has the superiority attributes. Using this algorithm the characteristics of 1st and 3rd clusters only found. And the other 3 clusters have the similar number of values.

No. Of Iterations: 7

Table 5.2 Farthest fast category table

Attribute	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Total	
Gender						0	
Std						0	
Home town	X			X		2	
School	X			X		2	
Income	X					1	

5.3 Filtered clusterer algorithm:

Using Filtered clusterer algorithm, 600 instances and 24 attributes were used, the data was analysed to find out the student's involvements in sports based on their attributes. Filtered clusterer algorithm the characteristics of 2nd ,3rd and 4th clusters only found. And the other 2 clusters have the similar number of values.

No. Of Iterations: 7

Table 5.3 Filtered cluster category table

Attributes	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total	
Gender			X	X		2	
Std						0	
Home town				X		1	
School				X		1	
Income					X	1	

From the comparative table it is concluded that the major correctly clusters were in the simpleKmeans algorithm. And the farthestfirst and the filtered cluster algorithm has the most similar values clusters. From the study it has been further concluded that the performance of the simplekmeans clustering provides more accurate result when compare to the other two algorithms namely Farthestfirst and filtered cluster algorithm.

6. CONCLUSION AND FUTUR WORK

The research study was conducted among the students of 9th to 12th std in the schools of rural and urban areas. Research was conducted with the structured questionnaire to collect the required primary data from the students collected data were analysed. The present study conducted among 600 students in and around kanchipuram within tamilnadu to identify the student's involvements in sports and games. Comparing india with other countries, even with small countries lagging behind them in sports and games. This thesis was conducted to identify the reason for this state of affairs to small extent. After analyzing the results of testing the algorithms and running them under different factors and situations, we can obtain the following conclusions: Among the extra curricular activities such as sports, drawing, music and dance the students are much interested in sports. Regarding the student's interest in sports 67% of students are playing games and the rest are not interested in playing games, mostly girls are not interested to playing games. The result of the analysation reveals that the students mostly interested in cricket next to cricket the students are interested in volleyball and football. The result of the analysis shows that the reason for their favourite sport is personal interest and favourite player. The results shows that the 46% of student's role model is their favourite player and 29% of student's role model is their friends. From this illustration its well known that most person get motivation from their friends and parents. 46% of student's motivate person is friends and 30% of student's motivate person is their parents. Most students wanted to participate in intra level competition compare to inter level competition. Generally 88% of students are not involved in sports club. The parents and the schools give pressure to the students for the achievement

in the field of education only. Next to this good health is not prevailed among most of the student due to non availability of nutritious food. 48% of students interested to concentrate on their studies. And 34% of students are having health problem.

Future work

There are many opportunities are available in obtaining job in private and government concerns for the athletes sports person. This should be given broad publicity to the student and the institutions. Those who are interested to further continue this project must pay attention in the following matter. The economic status of the parents and their occupation. The fund allotted for the sports activity in the institutions. The project may be conducted alone for the rural schools and urban schools. This project was done to the school level only. This may be extended to the college level. The participation of the girl is less than the boys. The reason and remedies for this may also be taken in to the consideration.

REFERENCES

[1] Bruce Kelley and Carl Carchia | ESPN The Magazine, (2013), The hidden demographics of youth sports.
[2] Matthias Chan Final Paper Independent Study (2010) A survey of sports in operational research.
[3] Robert P. Schumaker, Osama K. Solieman and Hsinchun Chen (2003) Sports Knowledge Management and Data Mining.
[4] S Schneider, B seither, S tonges (2005) Sports injuries: population based representative data on incidence, diagnosis, sequelae, and high risk groups.
[5] Pamela Wicker & Christoph Breuer (2005) Usefulness of Data Mining in sport management research.