**DATA MINING AND DATA WAREHOUSING**

**PROJECT**

**TOPIC:**

DATA MINING LARGE SCALE DATABASES USING GENETIC ALGORITHM.

**APPLICATION:**

FOOTBALL TEAM SELECTION PROCESS

**TEAM MEMBERS:**

ABHISHEK. V - 118014002

ASHWATH.P-118014010

HARIHARAN.K-118014032

**ABSTRACT**:

The process of selection of a team is a tedious process. A wide range of skills and ratings of each player must be considered. When given a large database consisting of each player’s attributes and personal information, it is quite difficult to filter out the best players. When an exhaustive search is used the execution time is quite long. As there are conflicting constraints and personal judgements involved, it is better to have a software for an optimized result. In this paper we will be giving a more effective way for approaching such team selection process by using genetic algorithm.

**INTRODUCTION:**

Football is a game played by two teams having 11 players each on a rectangular field with goalposts on each end. The offense, team with the ball tries to advance across the field by dribbling or passing the ball while the other team without the ball, defence tries to stop them by defending the ball from not getting past them and hence not scoring a goal.

**METHODOLOGY**

Genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution and is used for optimization and search problems to come up with solutions. A genetic algorithm is a part of evolutionary algorithms, which use techniques such as mutation, selection, crossover to generate solutions for optimization problems. The individuals of Genetic Algorithm are encoded in the form of strings where a collection of strings is termed population. Initially, a random population is created, which represents different points in the search space. An objective function is associated with each string that represents the degree of the goodness of the string. It is based on the principle of survival of the fittest where only some strings are selected after crossover and mutation to create the next population from the parent string. The process of selection, crossover and mutation continues for a fixed number of generations or till a termination condition is met. When the number of generations increases, so does the fitness rate.

The processes of genetic algorithm include:

1. Initialization

2. Selection

3. Crossover and mutation

4. Termination

**EXPERIMENT:**

Data of various football players were collected. This includes the player’s performance , skills and attributes . The database had information pertaining to 17,589 players. The top 10 teams were then filtered in which each team had 11 players. This was done by using a normal SQL(Structured Query Language) query. The selected teams were then randomly crossed over i.e. the players of the teams were crossed over to create a generation. Many such generations were consequently created. A mutation index was set up and if the crossed over player fitness doesn’t meet the mutation index value, the crossover is neglected and the selection process is carried out to the next generation. So, many generations were created to get the best 11 players from the 10 teams. These 11 players have the best possible attributes and skills. The experiment was then tested with the varying number of generations. The code was then optimized to reduce the time complexity and hence meet the objective.

**CONCLUSION:**

Fromthe above experiment, it can be clearly seen that the time complexity invariably reduces when genetic algorithm is used in the team selection process. This proves that the experiment done is a bit more efficient when compared to the exhaustive search mechanism.