CRICKET TEAM PREDICTION USING MACHINE LEARNING

Mr.B.Sakthivel, Abinaya A, Juraidha Rumman M, Madhuvarshini T, Tamilsudar B
Department of ECE
Vivekanandha college of Engineering for Women, Namakkal, India
sakthivel@vcew.ac.inabianbalagan592@gmail.comjuraidhavcew@gmail.com

ABSTRACT

Cricket player categorization is essential since it helps the coach and captain identify each player's position within the team and assign assignments accordingly. Players are categorised as batters, bowlers, batting all-rounders, bowling all-rounders, and wicketkeepers using performance data. By classifying individuals into five groups, this study seeks to accurately identify cricket teams in the one-day international format. The players are given an exceptional, very good, good, satisfactory, or bad rating based on both their past and present performance.This paper presents an improved model for the game of cricket, in which an eleven-person team is selected using a fair method. The performance, batting average, bowling average, opposing team strengths and weakness,etc..should all be taken into consideration while choosing players. To increase the precision of machine learning prediction models, feature optimization strategies that are inspired by nature are applied. The Combination of Cuckoo Search and Particle Swarm Optimization, known as CS-PSO, effectively incorporates the skills from both methodologies to produce trustworthy and appropriate solutions for successfully completing global optimization. Batter, Bowler, batting all-rounder, bowling all-rounder, and wicketkeeper selection accuracy was 97.14%, 97.04%, 97.28%, 97.29%, and 92.63%, respectively, using a mix of CS-PSO feature optimization and Support Vector Machine.

I.INTRODUCTION

To stay one step ahead of the competition and win over new fans or followers, cricket must innovate. One notable illustration of this is the One-Day International (ODI) format, which is arguably the biggest change in any team sport. In all versions of cricket, bowling and batting are the two key skills. Each cricket ball that is thrown generates a lot of data. A team's total performance is determined by evaluating and averaging each player's batting and bowling statistics. Cricket batsmen's performance is frequently evaluated using their batting average and strike rate, whereas bowlers' performance is generally evaluated using their bowling average, economy rate, and strike rate. The bulk of the scorecard's current criteria, meanwhile, are poor at identifying a player's intrinsic talent. For instance, batting average provides the typical amount of runs a batsman scores prior to losing his or her wicket. The ability of a player to score runs is determined by their batting average. However, it cannot determine how effective a batter is in scoring frequently. Similar to this, by examining the economy rate, one may determine a bowler's rate of run loss but not his capacity for wicket-taking. In order to measure cricketers' batting and bowling performances by using standard performance data, several performance metrics have been established. Success in ODIs depends on both the batting team's 50+ partnerships and boundary-hitting prowess, as well as the bowling team's dot-ball bowling and wicket-taking abilities. It is difficult to put together a team to play a certain competitor team since numerous factors, such as the strengths and limitations of both teams, must be taken into account. Picking the best players for each match in any sport is all that is required to predict how well the players will perform. In cricket, the exact number of 11 players is decided at the beginning of the play, and they play the entire game unless an injury happens. Based on prior

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.10

performances and other factors, it is necessary to estimate each person's performance and make a decision on whether or not they are a strong candidate to join the team. The roster was chosen after careful consideration of the balance of hitters, bowlers, and all-rounders. A wicketkeeper with outstanding figures behind the stumps and strong batting stats should have been on the team. Despite the appearance that fielding is an essential component of a play, bat and ball abilities are evaluated more highly than fielding.

II.EXISTING SYSTEM

Predicting the players' performance is nothing more than selecting the top players for every match in any sport. In cricket, precisely 11 players are chosen at the start of the play and remain fixed for the entire game unless an injury occurs. The individual's performance needs to be predicted with a choice as to whether the player is an exceptional contender for participation in the squad based on past records and other considerations. The decision for selection of the squad considered an enormous balance of batters, bowlers, and all-rounders. The team should have included a wicketkeeper with remarkable numbers behind wickets and impressive batting statistics. . The hybrid algorithm might utilize an assortment of two or more approaches to boost the algorithm's optimization capabilities. There are two types of hybrid optimization techniques. One direction is to employ a mechanism to pick one of the two optimization methods and then alternate between the two algorithms in the iterative optimization process. The other method enhanced the different techniques by using the primary formula of one algorithm. This method combined the two optimization algorithms' position updating formulae and chose alternative position updating formulas for optimization with a specific mechanism to apply to the hybrid system

III.**PROPOSED SYSTEM**

In this project, teams are chosen for one-day international cricket matches using machine learning algorithms. A balanced team is assembled by selecting batsmen, bowlers, batting all-rounders, bowling all-rounders, and wicketkeepers within the parameters of the regulations. The player's features pass through with the aid of the feature optimization algorithm, improving prediction accuracy dynamically. Methods drawn from nature are used to limit the number of input variables. Logistic Regression, Naive Bayes, K Nearest Neighbors, Support Vector Machine, Gradient Boosting Algorithm, Decision Tree, Random Forest, XGBoost, and CatBoost are nine classifiers that employ the chosen features from Nature Inspired techniques as input. To enhance the feature optimization efficiency for choosing pertinent features, a solo method is combined with the CS-PSO hybrid optimization approach. For this study, a machine learning framework was used to forecast the cricket team.

3.1 PARTICLE SWARM OPTIMIZATION (PSO)

The PSO method is a stochastic optimization technique based on swarms that resemble the social behaviour of animals including insects, fish, birds, cattle, and other creatures. These swarms employ a cooperative approach to food acquisition, with each swarm member adapting the search method in response to its and other members' growing knowledge. In order to satisfy the proximity and quality criteria, particles in PSO may modify their positions and velocities in reaction to changes in the environment. Additionally, the swarm seeks the ideal solution rather than restricting his movement with PSO. It is crucial to carefully select the particle population size N, the maximum number of repeats M, the inertia weight w, and other factors before applying the approach. The position and velocity of every particle are updated using the two equations below. inside the available solution area. Each person is treated as a particle in PSO and is described as a viable option to optimization problems in the solution space. It has the ability to memorise both the swarm's velocity and optimal locations. In order to change the movement of each dimension and determine the new position of the particle, each generation integrates the data from the particle. Due to its own experiences, the particle has trust in the direction and speed of its current state of motion and travels with inertia in that direction. The "social" element distinguishes between the particle's current situation and the swarm's optimal location globally (or locally). It mimics the movement of positive particles by taking use of social learning. With a greater inertia weight favouring global search and a lower inertia weight favouring local search, it is thought that inertia weight is utilised in 3.SO to balance out global and local search.

wvik + c1r1(pbk I xk I + c2r2(gbkxk I = ik+1 (1) Xik+1 = Xk I + Xk I + 1 (2) The location of a particle is represented by xi k in the equation above. The variables vik denote velocity, w denotes inertia weight, c1 and c2 denote learning factors, while r1 and r2 denote random integers with values between 0 and 1. Pbi k stands for the individual best of a particle, while gbk stands for the collective best of a swarm.

3.2 CUCKOO SEARCH (CS)

This strategy is based on the Levy flights' erratic movements and the brood parasitism of particular cuckoo species. Some cuckoo species lay their eggs in the nests of host birds while also destroying other eggs to increase the chance that their own will hatch. The eggs will develop into a full-grown cuckoo if the host birds do not find and destroy them. The ideal scenario would be for cuckoos to congregate and select the best place for breeding and reproduction as a result of cuckoo migration and environmental conditions. The host birds will either reject the eggs or quit their nests and start over if they discover the eggs are not their own. Usually, parasitic cuckoos select nests where the host bird has just deposited eggs. Compared to host eggs, cuckoo eggs grow a little bit faster. When the first cuckoo chick emerges, his natural inclination is to push the host eggs away from the nest. The cuckoo chick obtains a bigger portion of the food provided by its host bird as a result of this behaviour. Similar to this, a cuckoo chick may imitate its host chick's call to increase its chances of finding food. The cuckoo baby is the only child left in the nest after the host bird's offspring die of famine. The breeding habits of cuckoos may be used to several optimization problems. Each egg in a nest, including cuckoo eggs, denotes a certain type of solution. The goal is to replace less-than-ideal solutions in the nests with new and maybe superior ones. Selection of the fittest and environmental adaptability are two crucial characteristics. These may be realistically translated into intensity and variety, two essential features of modern metaheuristics. While intensification tries to search the best current solutions and choose the best candidates or solutions, diversification ensures that the algorithm can effectively explore the solution space. The Levy flying technique stated as an equation leads to the next answer.

xi t+1=xt I + Levy (β) (3) (3) The future state xi t+1 of the levy flight global randomised walking technique is simply defined by the current state xi t and the transition probability levy (). A Levy flight's random step size is computed using the Levy distribution's infinite variance and mean. Levy = t * 1 * (0 * 2 *)

IV. HYBRID PARTICLE SWARM OPTIMIZATION AND CUCKOO SEARCH ALGORITHM

Any population-based algorithm must seek and use data in order to function successfully, which is a well-known fact. The PSO algorithm can easily enter the local optimal solution. On the other hand, the random walk method of the cuckoo algorithm can increase the diversity of the solutions in the search space. As a consequence, this research recommends combining the PSO algorithm with the random walk methodology to produce a novel hybrid optimization technique. The goal of this hybridization technique is to choose the best teams for one-day international matches by extracting pertinent qualities from the data of batsmen, bowlers, allrounders, and wicketkeepers. To enhance the search for global optimal solutions, the Levy flight approach is used in place of the random values r1 and r2. The phases of the hybrid optimization method CS-PSO.

1. Algorithm
2. Set the cuckoo search parameters in motion.
3. Separate the populace into several categories.
4. Utilizing the fitness value of each person, apply the cuckoo search method to identify the local best solution for feature optimization.
5. Use the PSO algorithm's input population as the local optimal solution discovered by cuckoo search.
6. Begin by initialising the particle from the PSO input population.
7. Formulas 1 and 3 are incorporated into the PSO algorithm to facilitate the effective search for optimal solutions.
8. (c1 Levy ()) (pbk I xk I + (c2 Levy ()) (gbkxk I = wvik + (c1 Levy ()) (5)
9. It aids Particles in the PSO algorithm in effectively obtaining global solutions. because levy flight improves the local walk of particles.

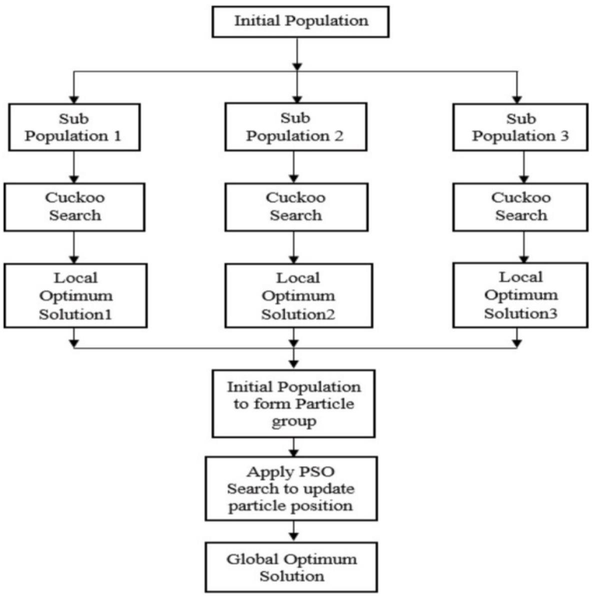
ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.10

10. An optimal output of the optimization process is the best result from the CS-PSO hybrid strategy.

The feature optimization methods use the features from Tables 1 and 2 as input and eliminate the characteristics that are unnecessary. Feature optimization techniques are used to identify the factors that have the greatest influence on determining a player's strength. Machine learning classifiers use the improved features that result from feature optimization techniques as input. After that, one of five classes is selected by machine learning algorithms to categorise players.

V.LEARNINGALGORITHMS/MODEL SELECTION

This study employs Logistic Regression, Nave Bayes, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting method, XGBoost, and CatBoost algorithms. Because of the classification problem, these algorithms are chosen. The winner.



PSO-CS Hybrid approach for feature optimization

The formula for the prediction classification issue is y = f(x), where x seems to be either a single or a collection of independent variables, and y is the dependent variable.

VI.ALGORITHM1BATSMENSTRENGTH

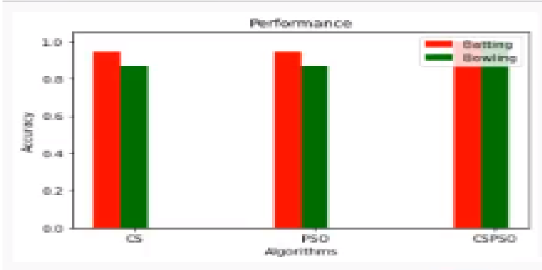
1: for all players p do
2:α_batsmen_score=0.30*run_scored+0.05*not out_innigs+ 0.20*bat_avg + 0.15*bat_sr+0.15*milestone_reaching_ability+ 0.10*no_of_4's_6's+0.05*high_score- 0.05*no_of_zeroes
3:β_pos_score=max (α_batsmen_score_at_each_position)
4:γ_inningwise_score=0.40*α_batsmen_score_first_inning+0.60*α_batsmen_score_seocnd_inning
5:x_venue=0.35*α_batsmen_score_home_matches+0.65*α_batsmen_score_away_matches
6:y_opponent=0.70*α_batsmen_score_strong_opponent+0.30*α_batsmen_score_weak_opponent
7:w_yearwise=0.20*α_batsmen_score_current_year+0.80*α_batsmen_score_last_five_year
8:batting_score=0.25*α_batsmen_score+0.10*β_pos_score+0.15*γ_inningwise_score+0.10*x_venue+0.15*y_opponent+0.20*w_yearwise+0.05*captain
9: end

Classifier	Accuracy								
	WithoutFeatureOptimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
LogisticRegression	80.64	84.28	83.43	82.31	86.04	86.91	87.79	91.56	91.86
NaïveBayes	77.85	79.54	82.85	81.21	86.64	91.08	90.35	90.57	92.76
KNN	78.92	80.30	87.87	89.87	89.97	89.61	90.56	91.87	93.89
SVM	87.09	89.28	88.29	89.31	90.33	90.74	93.46	94.33	97.04
DecisionTree	79.28	82.35	85.17	84.02	88.42	89.62	90.39	92.36	92.71
RandomForest	86.62	87.53	87.34	88.60	89.57	90.64	90.85	95.53	95.92
GBM	85.07	86.25	88.17	89.33	88.37	89.24	91.27	94.47	96.05
XGBoost	83.28	84.47	85.49	85.24	89.07	92.53	93.12	93.73	95.67
CatBoost	85.46	86.33	87.02	90.69	90.64	91.64	91.34	93.63	96.22

VII.RESULT AND DISSCUSION

In order to predict the team for an ODI match with the highest degree of accuracy, this work combines an ensemble of machine learning models with feature optimization techniques. The feature optimization techniques produce machine learning models with excellent accuracy using fewer input features. The qualities that have the greatest impact on player selection are chosen using feature optimization algorithms. The Natural Design The metaheuristic method efficiently chooses feature subsets from a dataset and is motivated by the behaviour of natural creatures or swarms. Not all features influence player evaluation equally. Some traits, as opposed to others, have a stronger impact on the outcome of the machine learning classifier. Approaches to feature optimization locate traits with greater weights to improve the classifier. To choose the team for one-day international matches, we use a variety of Nature Inspired algorithms using a hybrid CS-PSO system. Since they describe a batsman's consistency and scoring capacity, batting average, strike rate, and milestone-achieving ability are crucial factors in batsman selection. Bowlers' success in away matches, bowling average, strike rate, and economy rate are all important factors. Batting strength-related factors like batting average and strike rate have a positive influence on the choice of batting all-rounders, whereas bowling features have an impact on the choice of bowling all-rounders. In comparison to batting features, the number of catches behind the wicket and stumpings significantly influences the wicketkeeper choice. The batting prowess of the wicketkeeper is less significant than his glub. The ideal combination of machine learning model and feature optimization approach is found after repeatedly estimating the model on training and testing data from the dataset. It is better to choose a team using a hybrid CS-PSO and SVM algorithm. With less training time, models' accuracy is improved using all other methods for feature optimization. The CS-PSO hybrid approach and SVM algorithm are used to select the players with high ratings and scores. Only the selection member will take these players' abilities into account when deciding whether to include them on the team. We contrast the team composition with our strategy with the squads chosen for the India-Australia Series [28] and the India-New Zealand Series, both of which are depicted in Table 10. Table 10 shows the performance of the cricketers who were actually chosen for the team and the players with various levels of experience who were passed over for a spot on the national team. Ajinkya Rahane and Bhuvneshwar Kumar were not chosen for the

India-Australia Series despite having strong player scores and ratings. Ajinkya Rahane and Ravindra Jadeja did not play in the India-New Zealand series despite having strong records and chances to make the team. However, the majority of the players chosen from our models, based on their skills, are included in the team playing eleven in the India-Australia and India-New Zealand series.



PERFORMANCE

VIII.CONCLUSION AND FUTURE WORK

For institutions looking to increase performance, selecting the right personnel is a crucial challenge. One example of this issue is team selection, where the objective is to determine team members. The majority of the time, candidate ratings are applied to better both the individual and team hiring procedures. The right player choice for each match has a significant influence on the outcome of the game. Based on a reasonable projection of the number of runs a batter will score and the number of wickets a bowler will take in a match, team organisation members may choose the best players for each game. However, these conclusions can only be made with information obtained from a variety of sources. In this study, we developed a model for choosing an 11-member team based on the data and characteristics of the players. This article looked at the several player classifications used in one-day international cricket. Using a dataset of 414 Indian ODI players who were playing under a rule constraint for squad selection, we investigated classifying players into one of the five categories using machine learning techniques and a feature optimization method. In this study, we employed nine machine learning techniques to identify the class into which each player should be placed. Following the initial application of the aforementioned techniques, SVM's prediction accuracy for choosing batters is 93.54% and for choosing batting all-rounders is 87.29%, respectively. The accuracy of bowlers using SVM is 87.09%, that of bowling all-rounders is 85.70%, and that of wicketkeeper selection is 84.21%. We improved the forecast accuracy even further with the appropriate parameter choices. Using a combination of SVM and CS-PSO, we increased the prediction accuracy for batsman selection up to 97.14%. Similar tothis, using a CS-PSO improved the SVM's accuracy for selecting bowlers and bowling all-rounders to 97.04% and 97.29%, respectively. For batting all-rounders, the SVM with a CS-PSO offers a maximum prediction accuracy of 97.28%. By choosing appropriate values for the parameters, the hybrid technique identifies the wicketkeeper with an accuracy of 92.63%. Our findings concluded that the Nature Inspired algorithm significantly outperformed machine learning methods. The study's findings are useful to both cricket officials and players in a number of ways. Teams' coaches, captains, and player selection committees can utilize these findings to identify the best players. Other factors that impact the performance of the player may be added to this study in the future. Future studies should incorporate more performance data; such as details on the teams that the athletes compete against. The final objective will be to improve the classification model's accuracy after adding more features to the data. This approach may be used to predict the winning side in Twenty 20 matches with the right data and feature adjustments.

IX.REFERENCES

[1] Ishi MS and Patil JB a study of artificial intelligence techniques for cricket team selection and result prediction. Networks Syst Lect Notes 2021; 173. https://doi.org/10.1007/978-981-33-4305-4 12 LNNS:143-56

[2] Ishi MS and Patil JB. a research on the effects of team makeup and the ideal variables needed to forecast the outcome of a cricket match. 2020;100:389-99. Lect Notes Networks Syst. https://doi.org/10.1007/978-981-15-2071-6 32 .

[3] Moth-flame optimization algorithm: a unique, nature-inspired heuristic paradigm. Mirjalili S. Knowledge Base Systems (2015) 89(7): 228–49. DOI: 10.1016/j.knosys.2015.07.006 [Internet].

[4] Firefly method, stochastic test functions, and design optimization Yang XS. 2010;2(2):78-84 Int J Bio-Inspired Comput. The citation is 10.1504/IJBIC.2010.032124.

[5] The Whale optimization technique by Lewis A. and Mirjalili S. 2016;95:51–67. Adv Eng Software https://doi.org/10.1016/j.advengsoft.2016.01.008 [Internet].

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-10, ISSUE-4, 2023
DOI: 10.21276/ijcesr.2023.10.4.10