International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Tennis Match Prediction using Machine Learning

Manjeet Gupta[1], Ranjeet Gupta[1], Salman Ansari[1], Mrs. Darshana N. Tambe[2]

[1]Student, Dept. of I.T. ,VPPCOE & VA,Mumbai University ,Mumbai, India
[2]Professor, Dept. of I.T. ,VPPCOE & VA,Mumbai University ,Mumbai, India

ABSTRACT

Winning is the primary goal of any sport. Predicting the winner of the match before has gained loads of attention from sports organizations and potential bidders because it involves loads of your time and cash invested within it. Nowadays, sports organizations notice the worth knowledge of information} and therefore the science within the data which might be used as a plus to players coaches conjointly the potential bidder's victimization machine learning techniques. lawn tennis could be a difficult and unpredictable sport, nevertheless the foremost exciting sport that is enjoyed by fans from everywhere on the planet. Machine learning techniques are helping us to predict the outcomes of lawn tennis matches victimization varied attributes. the most objective of this project is to predict the winner of the match victimization individual player statistics and with the assistance of assorted parameters of lawn tennis serve and therefore the individual set score of each player, the winner of every match is foreseen. The point-by-point provides insight and a higher understanding of the sport. The task is achieved with varying degrees of success through the implementation of a Neural network.

Keywords—Prediction, Performance, Education, Marks, Machine Learning, Linear Regression Algorithm, Dataset, Evaluation, Data Preparation, feature selection, Libraries, Results.

I. INTRODUCTION

lawn tennis is an extremely fashionable sport that is enjoyed and worshiped by fans from everywhere on the planet. lawn tennis has four major tournaments referred to as the sweep tournaments specifically the Wimbledon, Australian Open, United States Open, and French Open.

it's sometimes contended by players on 3 differing kinds of surfaces (Clay, Hard, Grass) . lawn tennis is an especially unpredictable sport that is contended by sports players from numerous backgrounds and different design. every player contains a distinctive vogue and technique that makes the sport even a lot of attention-grabbing and difficult to predict the winner. Today, machine learning is employed in several sports like association football, cricket, baseball, lawn tennis. AS we all know information is all over and lawn tennis is outlined by information, and machine learning techniques area unit already creating waves within the field of lawn tennis not just for skilled players but conjointly for coaches, fans, and potential bidders.

The purpose of prediction in a table tennis game is to predict the result of future matches consistent with the present status of skills and tactics. It's important to rearrange exercises properly and take some countermeasures during competition. Particularly within the team competition, we will arrange the playing order to support the player's abilities. .

II. LITERATURE SURVEY

Machine Learning for Professional Tennis Match Prediction (Andrecornman-2018) (IEEE)
Algorithms used: SVM, Logistic Regression, Random Forest, Neural Network. Data Set used TENNIS UCLdataset. AdvantagesThe Neural network model has higher accuracy than any other model, which is 70%. DisadvantagesIt predicts only higher-ranked players because the prediction is based on the ranking. They have used fewer attributes for prediction. This model does not capture the low-ranked players.

Decision Tree and the Artificial Neural Networks to Predict the Outcome of Table Tennis Matches
(IEEE) (Jie WANG -2019)
Algorithms used: Decision Tree, Neural network. Data Set is used from the Australian Opendataset. AdvantagesAccuracy of the neural network is higher than the Decision tree algorithm. Disadvantage It takes a higher time to train the model. They have used fewer attributes such as skills and tactics only.This model requires high computational power.It is only for the Australian Open series.

Predicting the Winner of a Tennis Match Using Machine Learning Techniques
(IEEE)(Akshaya Sekar -2019)
Algorithm used: SVM, Logistic Regression, Naïve Bayes, Random Forest. Dataset is used from Us Open 2014 dataset.Advantages SVM outperformed another model.Disadvantage This model only predicts the matches in us open and the dataset they have used is old.The prediction is only for us open .

Predicting the outcomes of tennis matches using a low-level point model

(DEMETRIS SPANIAS-2019)

(IEEE)

Algorithms used: Generic point model. Data Set used from ATP tournament dataset. AdvantagesThis model is better than the existing Model. DisadvantagesThe prediction of this model is 67% which is less than expected. It gives the average probability of the predicted winner.

Disadvantages: Comparatively smaller data set making predictions a bit unreliable.

Tennis Match Prediction Model

(IEEE)(Jiali pan-2018)

Algorithms used: SVM, Logistic Regression, Random Forest, Neural Network. Data Set used Davis Cup dataset. The advantagesTraining process is fast and it takes less time to give the prediction. DisadvantagesThe overall accuracy of this model is 62 % which is less than expected. This model only predict the winner of the tournament

III. PROPOSED SYSTEM

In the above section we discussed how Machine Learning can be used to predict the performance of the players and identifying the weakness as early as possible so appropriate actions can be taken to enhance their performance. This paper is aimed to predict the outcome of tennis matches and make betting better for many sport lovers.

Softwareused Tech: Python, Machine Learning. Libraries: NumPy, pandas, matplotlib, seaborn, scikitlearn. Implementation: Jupyter Notebook. Website: HTML, CSS, JavaScript, Flask.

Hardwareused Processor: i3 4th Generation or higher, 4GB RAM, 1GB Storage.

Datasetused: The dataset contain attribute of date, series,court,surface,Best of,Winner,Losser, Winner Rank,Loser Rank,Winner points ,Losser Points ,round 1 ,round2 ,round 3 points ,serving points,break points .These are some essential attribute for this project.

The entire dataset was extensively analysed andfeatures

(attributes) were tested against the label (final attribute) using the dataset was splinted in 2 parts to avoid overfitting for training, testing and validating sets, we split the dataset into 75:25 for training to testing sets ratio. From this, 75% of the dataset is used for training setsTo get the right predictions, we must construct the data set and transform the data correctly.The first step is collecting the data from the various data sources. In our case, the data has been collected using a betting website There are a lot of data about tennis.

- o The Second step is Data Preprocessing ,in this process we clean the data to fed our machine learning algorithm..It makes data more comfortable.
- o The third step is Feature Extraction,In this process
- o In the fourth step, we divide the clean data into training and testing dataset, is fed to the Machine Learning algorithm .
- o By Using Machine Learning Algorithm we get the accuracy of training dataset and testing dataset and after we will analyze our result.
- o At the end ,the Machine learning algorithm gives the. trained model which will predict the output of matches.

Data Collection : We have used various attribute to get the accurate prediction of the matches and makes more easier to understand .We have taken dataset from 2000 to 2021 tennis matches of mens and womens.Our main purpose is to predict the outcome of us open, Wimbledon ,French Open,Aus Open matches so we will take data of this tournament.

Attribute Information :

ATP, Location,Date,court,Surface,Round,best of,Winner ,Losser,Winner Rank,Losser Rank,Winner points,Losser points, Winner round1 points ,losser round1 points ,Winner Round2 points,Losser Round2 points,Winner Round3 points ,Losser Round 3 points ,Winner Round 4 points ,Losser Round 4 points ,Winner Round 5 points ,Losser Round 5 points,Winner's total winning sets, Losser's total winning sets,Breaking points ,Service point.

Data Preprocessing :

In data Preprocessing we will clear the data to get normalize data which will important for our machine learning model.We will remove unnecessary attribute from the dataset.

Feature Extraction :

In Feature Extraction we have calculated the winning percentage of each player on different surface , their last 60 week performance on grass, clay .Here we have also calculated the winning percentage of each player on every round and how their performance is.

Feature divide:

After Feature Extraction we have divided the feature into two parts 75 % for training and 25 % for testing.

Machine Learning Algorithm :

In this paper we have used three algorithm neural network, Linear Regression and Random Forest.

LinearRegression: Linear Regression is a machine learning algorithm i.e. based on supervised learning. It performs regression task. A regression models a target prediction value based on independent variables provided. It is mostly used to find out the relationship between variables and forecasting.

Random Forest: Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

Neural Network:The human brain is composed of 86 billion nerve cells called neurons. They are connected to other thousand cells by Axons. Stimuli from external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural network. A neuron can then send the message to other neuron to handle the issue or does not send it forward.

IV. RESULTS

In order to test how well our models relate to reality, we have programmed back-testing software that replicates tennis matches between players. This software uses statistics, based on information which would have been available at the time of the match, to calculate, using the techniques outlined in this paper, the probability of a server winning a point and subsequently calculate the probability of winning the match in a hierarchical fashion. The back-testing software executes this process for a wide set of matches in order to analyse how the models perform by comparing the predicted results with the reallife results. For each match modelled, the player that has a probability >0.5 of winning the match is considered as the predicted winner of the match.

After all the process we found these features are important are for our predictions :

- Court : outdoor or indoor place where the tennis match is happening.
- Surface : The surface in which players are playing like clay ,grass.
- Ranking : The ranking of player is also important feature ,it shows how player consistent.
- First Serve : First serve win percentage define the most important aspect of the match.
- Breaking points : Breaking points is most also decide the game win percentage.
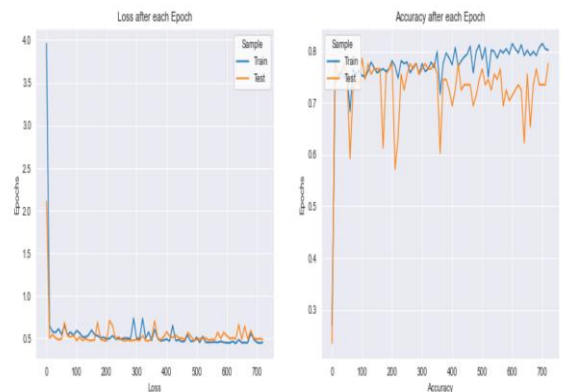
Model	Accuracy	F1 Score
Neural Network	0.82	0.86
Linear Regression	0.77	0.73
Random Forest	0.72	0.71

So , on the basis Accuracy of every model , Neural Network is best model for prediction .

V. CONCLUSION

This research study shows that by using different variables we can predict the outcome of tennis matches with higher accuracy. We have used dataset of tennis matches from 2000 to 2021.In this study found that on the basis of surface we can predict the outcome of matches with higher accuracy. In this paper we have used various machine learning algorithm like Linear Regression, Random Forest , Neural Network . Neural Network has Higher accuracy than any other model so neural network is best model for prediction. The main objective of this paper is to predict the outcome of tennis matches by using different attribute .

This research could be further enhanced by selecting the data which helps to predict the winner of the match based on:- Weather Condition: The weather condition is one important aspect to consider in predicting the winner of a tennis match. Getting the weather data on the same day of the match will help to predict if it has any influence on the match results.

In Play bets: Various In play bets can be conducted with the appropriate data, such as bets on predicting what the next shot will be or how many aces will the player hit before winning the match will be interesting and challenging to predict in the future.

Intensity of previous matches.

REFERENCES

Haghighat, M., Rastegari, H. and Nourafza, N. (2013). A review of data mining techniques for result prediction in sports, Advances in Computer Science: an International Journal 2(5): 7–12.

Hassanniakalager, A. and Newall, P. (n.d.). A machine learning perspective on responsible gambling. 2018.

 Learning, M. (2017). Final project report: Real time tennis match prediction using machine learning.

Mantovani, R. G., Rossi, A. L., Vanschoren, J., Bischl, B. and De Carvalho, A. C. (2015). Effectiveness of random search in svm hyper-parameter tuning, 2015 International Joint Conference on Neural Networks (IJCNN), Ieee, pp. 1–8.

Pathak, N. and Wadhwa, H. (2016). Applications of modern classification techniques to predict the outcome of odi cricket, Procedia Computer Science 87: 55–60.

Philpott, A. B., Henderson, S. G. and Teirney, D. (2004). A simulation model for predicting yacht match race outcomes, Operations Research 52(1): 1–16.

Prasetio, D. et al. (2016). Predicting football match results with logistic regression, 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), IEEE, pp. 1–5.

Sipko, M. and Knottenbelt, W. (2015). Machine learning for the prediction of professional tennis matches, MEng computing-final year project, Imperial College London.