Super Predictor Of Indian Premier League (IPL)

project report

INTRODUCTION:

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

As the process of analyzing raw data to find trends and answer questions, the definition of data analytics captures its broad scope of the field. However, it includes many techniques with many different goals.There is no doubt in higher the collection of data, more will be the effect of analysis. Hence there is a requirement of proper arrangements of data which can handle a large amount of information in a clear-cut view which should be comprehensible to all.

Worldwide various organizations are working on this technology and it is first setup in early 1962 when mathematician John. W. Tukey predicted the effect of modern-day-electronic computing on data analysis as an empirical science.

Fake news detection, Road lane line detection, sentiment analysis detecting , Parkinson's disease color detection with python, brain tumor detection with data science. Leaf disease detection are some of the very popular data science projects worldwide and cricket data science and data analysis is one among them.

When it comes to developing teams, this system is beneficial. Each team whether superior or super growing can implement these intelligent technologies it will be of great benefit and last at not the least it is more reliable than human prediction since we all know the fact data don't lie.

OVER VIEW:

Data analysis is one of the best methods to simplify or minimize difficulties in getting data. The main aim of data analysis is to achieving data efficiency, reducing traffic congestion, to get a clear view of information that we are seeking, reducing our time, providing more effective data, gives better ideas about their levels and is much comfortable and easy to view over data we want.in this IPL model, It covers all sorts of information and some of them are number of times a team won the match, number of times a team lost the match, runs of each team and players and many other information

The overall process of model is to collect data, analysis of the Cricket over a wide range and provide a clear view of all the required data.

PURPOSE:

Compared to other sports, cricket is one among few sports which generates high amount of data.

These raw data may seem useless because of large amount of data.

But these data are gem of immense value as higher the amount of data, higher the accuracy is.

Visualizing such raw data using IBM Cognos Analytics and gaining outputs according to our need

Visualization gives you answers to questions you didn't know you had
The goal is to turn data into information and information into insight.

LITERATURE SURVEY:

Indian Premier League (IPL) is one of the more popular cricket tournament, and its financial strength is increasing each season, its viewership has increased markedly and the sad part is even betting market for IPL is growing significantly every year. With cricket being a very dynamic game, bettors and bookies are incentivized to bet on the match results because it is a game that changes ball-by-ball. There are some papers which investigates machine learning technology to deal with the problem of predicting cricket match results based on historical match data of the IPL. Influential features of the dataset have been identified using filter-based methods including Correlation-based Feature Selection, Information Gain (IG), Relief and Wrapper. More importantly, machine learning techniques including Naïve Bayes, Random Forest, K-Nearest neighbor(KNN) and Model Trees (classification via regression) have been adopted to generate predictive models from distinctive feature sets derived by the filter-based methods. Two featured subsets were formulated, one based on home team advantage and other based on Toss decision. Selected machine learning techniques were applied on both feature sets to determine a predictive model. Experimental tests show that tree-based models particularly Random Forest performed better in terms of accuracy, precision and recall metrics when compared to probabilistic and statistical models. However, on the Toss featured subset, none of the considered machine learning algorithms performed well in producing accurate predictive models.

Cricket is unique in that three formats of the game, of varying duration and intensity are played at elite level; Test, One Day and Twenty20 (T20) cricket. The wicket-keeper is a vital member of the fielding side; the only player who could be involved with every delivery of the innings. Despite this, there is a paucity of literature on wicket-keeping, with most research having focused on fast bowling and batting. some review are primarily concerned with identifying the performance demands associated with wicket-keeping and the conditioning of these cricketers. To appreciate the performance demands of wicket-keeping, an inte-grated approach has been taken where the mental, technical, physiological, physical and tactical requirements of wicket-keeping are discussed. From this information, recommendations are made and

suggestions for future research detailed.

EXISTING PROBLEM:

There are no well proven analysis model for cricket

PROPOSED SOLUTION:

- Teams won most number of matches
- Teams lost most number of matches
- Teams with highest batting score
- Teams with highest wickets
- Teams with most win while batting first
- Teams with most win while fielding first
- List of matches won by each team in each season
- Winner by toss winner
- Matches won by runs
- Players with most player of the match awards
- Cities that hosted the maximum number of IPL matches
- Top run scorers in IPL
- Top wicket takers in IPL
- Top fielders in IPL
- Most winning teams for each season
- On-field umpires with maximum number of matches
- biggest victories while defending a high total
- Biggest victories while chasing a high total
- Top players with highest runs in single season
- Top players with highest wickets in single season

THEORITICAL ANALYSIS:

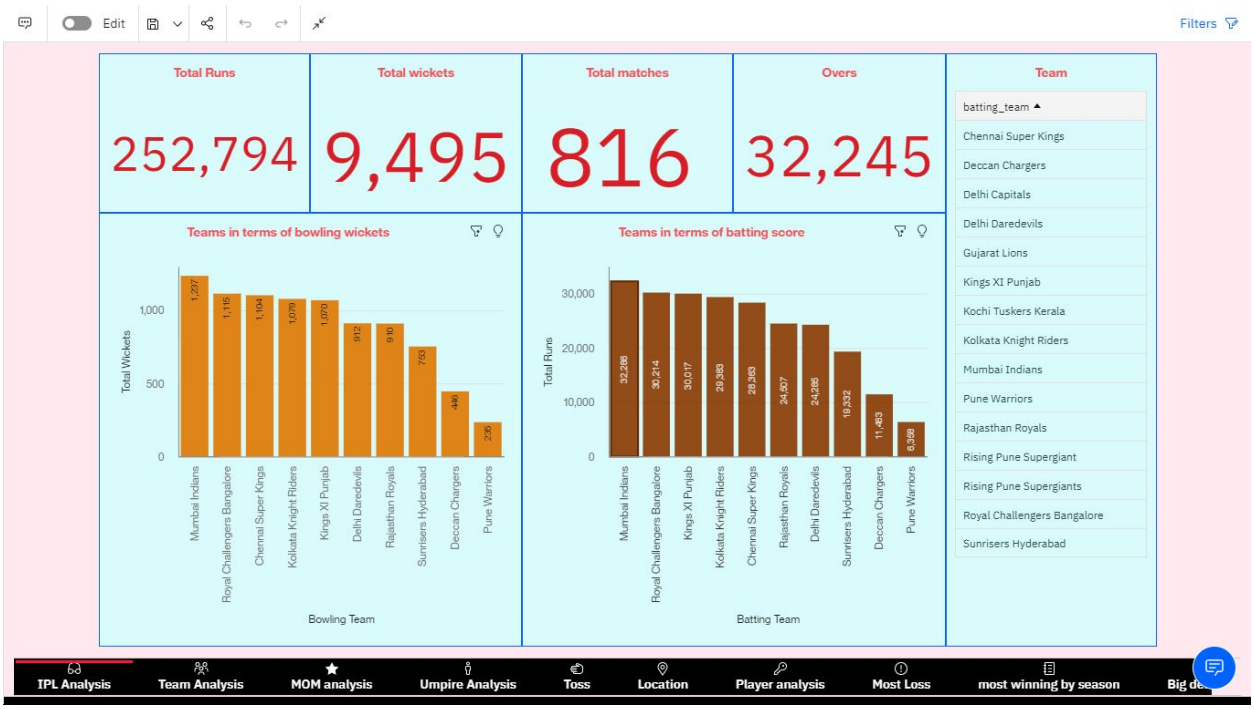
I had provided few of the tabs of my dashboard for the theoretical analysis

for the first tab, I had provided name as IPL analysis

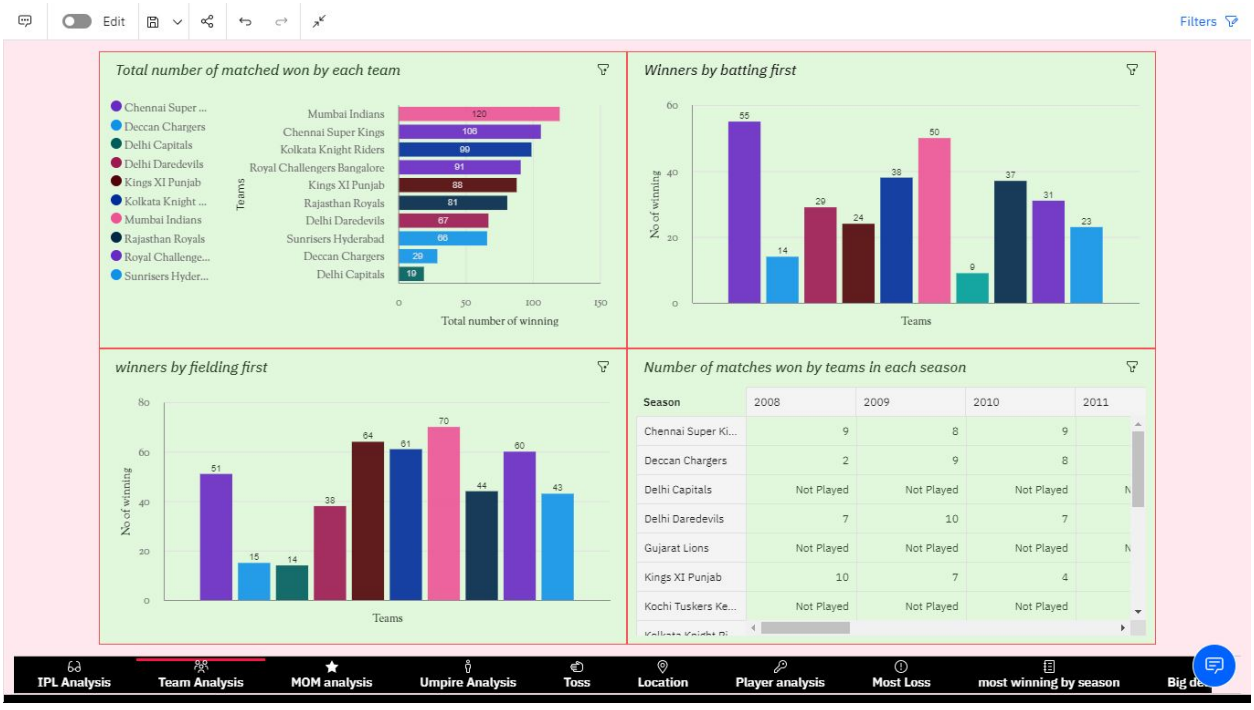
It displays the list of teams participating, total runs, total wickets, total matches, total overs

whenever a particular field is selected, the visualization gets displayed for that particular selection

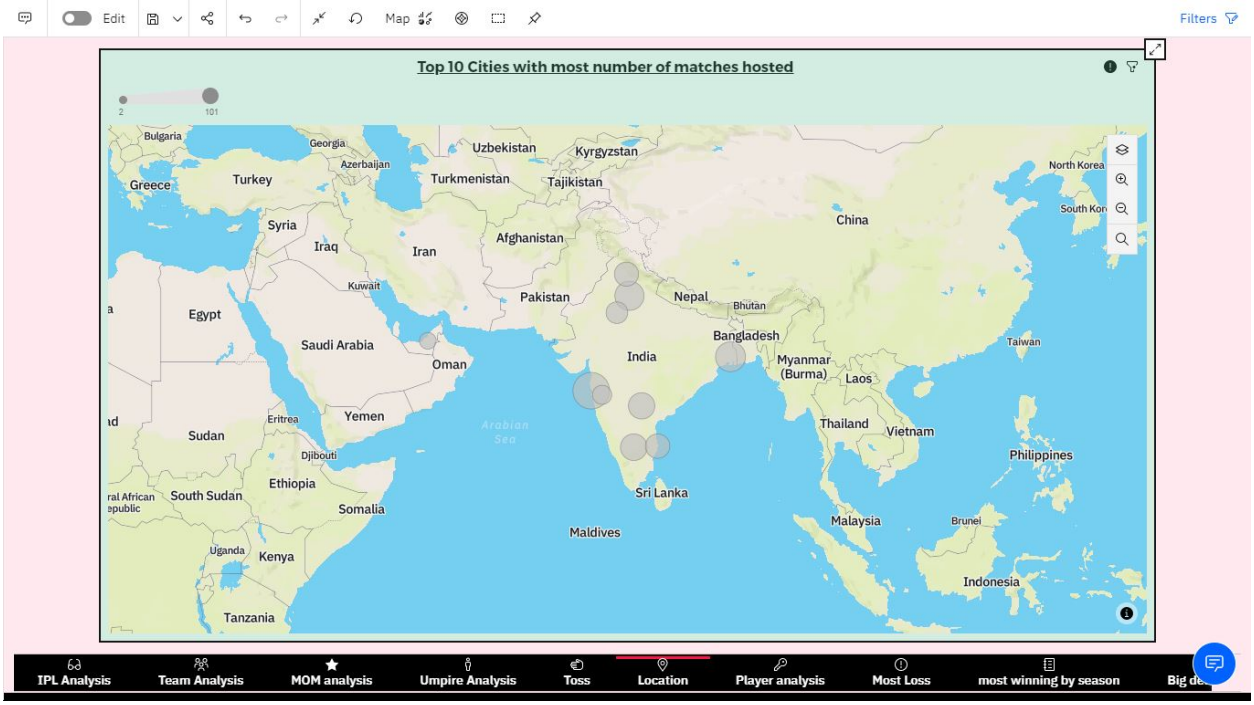
visualization occurs smoothly



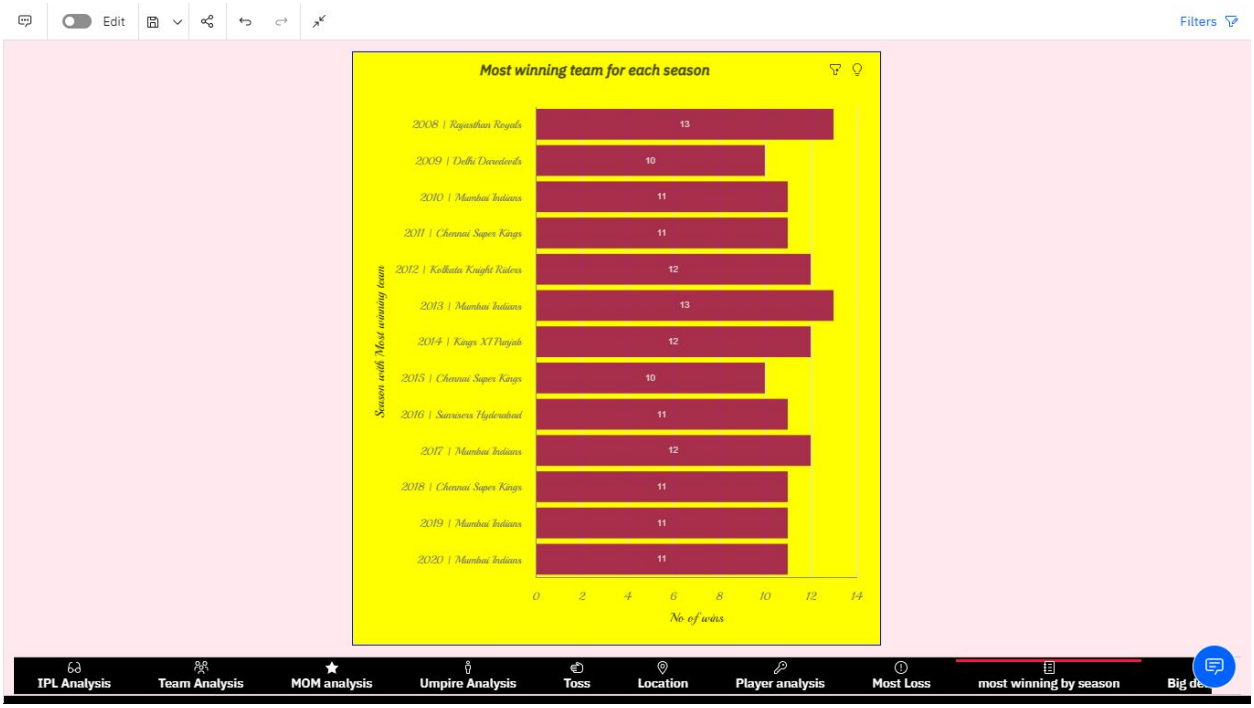
In this second tab named as team analysis
It displays Total number of matches won, Total number of matches won while batting first, Total number of matches won by each team in each season



In this tab named as Location
It displays the top 10 cities which hosted most number of matches



In this tab names as most winning by season
It displays Top team which won most number of matches



BLOCK DIAGRAM:

data

IBM Cognos Analytics

Output

HARDWARE/SOFTWARE DESIGNING:

No hardware/software was designed

EXPERIMENTAL INVESTIGATIONS:

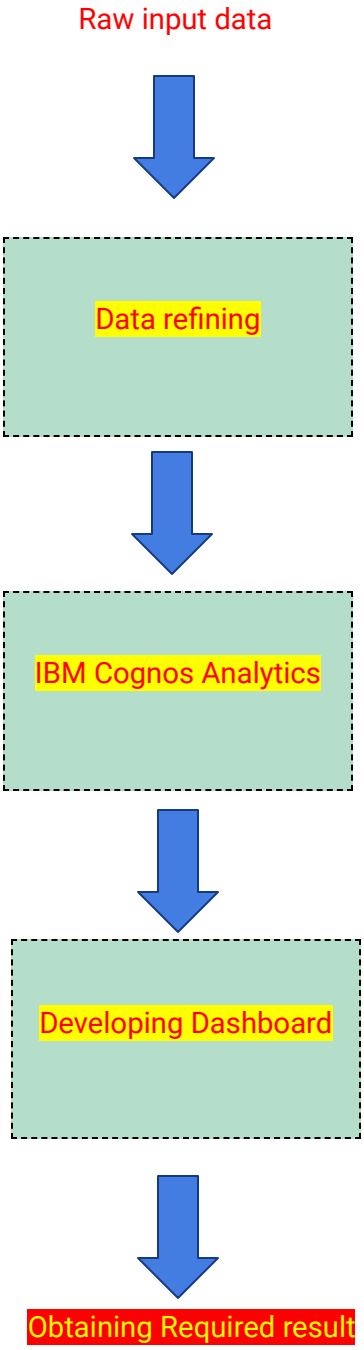
Many calculation fields been created for investigating in dashboards

some of the calculation fields were:

- Total number of matches
- Total number of overs
- Total number of winning
- Host number of matches
- season

the season field was created using function known as year which extracts only the year from data field

FLOW CHART:


Raw input data
Data refining
IBM Cognos Analytics
Developing Dashboard
Obtaining Required result

RESULT:

Data science is the foremost developing system. It gives a clear view over any information we seek for, having reports on available platforms develops their performance
sports play a vital role in database. especially cricket has a huge impact over it. It makes more easy to maintain a team as well as their performance. since anything about their performance are clearly mentioned, the players induce themselves to perform better and they try to make a good game.
This makes the people to change in a positive manner and to develop their self discipline in their professionals

ADVANTAGES :

1. According to the rules IPL [Indian Premier League] Council and BCCI, a team franchise would be allowed to retain a maximum of four players only. There is also a condition that out of those four retained players, two must be Indian players. So as a result, many well recognized players will be on bid. So, at mega auctions, many players will be on bid including some well recognized players.
   At such condition, this IPL predictor analysis model will be much useful for team franchises and would be a, eye opener for them to bid on some players based on their past performances and statistics.
2. This IPL predictor analysis model will be much useful for broadcasters to provide statistics, comparisons of team or individual players and always keeping the audience engaged.

3. This IPL predictor analysis model will also be useful for sponsors to opt which team franchise based on their performances and statistics, so that the sponsor's brand or product reaches to as many people as their expectation.

DISADVANTAGES:

It acts as key of success for many online fantasy games where many youngsters are wasting and losing money and time in that

At times,few players performs beyond expectations and at such situation, the prediction fails

APPLICATIONS:

Data Science is not only helpful for predicting the most favorable team for the tournaments but also helps generate valuable insights for other aspects.

Team management and leadership

Like any other team game, in Cricket too, it matters a lot to understand the dynamics of how each player performs during the match, particular unique strengths that can be leveraged against a team. By analyzing the data about the players' performances and drawing insights the captain can decide on the batting and bowling order. This will help in bringing in scientific approach to get the best out of the team and demonstrate well-informed leadership on and off the ground.

Performance improvement and preparation

Since every player's performance can be studied to analyze their strengths, weaknesses and unique attributes, there is huge scope for trainers and coaches to make use of the insights in improving the outcomes. Through strategic application of Data science they can strive for consistent performance by preparing for various challenges with different teams in varied circumstances.

Insightful prediction of results

As an audience, the question of which team is going to win keeps us on the edge throughout a match. With the use of Data Science, it is not just guess work anymore! The same engages several others who are connected with the game – the organizers, authorities, players, media and the entire nation for whom the game is nothing less than a religion.

Connect with enthusiasts and fans

With billions of enthusiasts and fans, Cricket has huge potential to connect with them in innovative ways. Data about their favorite players will definitely excite the fans and they will keep asking for more. Many serious enthusiasts look for informed insights to learn more about how the game is being played under different circumstances and locations. Then, the insights thrown by data also opens up new avenues and innovative options to sponsor the game and

connect with the admirers.

Keeps cricket fans engaged

The statistical data related to a single batsman and bowler highlights the wickets left, the way the ball was swung, runs scored per deliveries faced, the way each player responded to the delivery, and so on. This data allows fans to understand the game in depth rather than just looking at the match proceedings.

Help captains make the right decisions

Data analytics can help solve the uncertainty attached to a bowler or a batsman's average performance. What's critical is to know how they will perform in a given circumstance. Collectively, all of this data has the potential to create vast opportunities to analyze and draw meaningful insights, which then help predict or classify future events. This, in turn, helps captains make the right decisions, on and off the field.

CONCLUSION:

Super predictor of IPL will definitely going to affect the game in a positive way. It will reduce time of discussion and planning and helps in enhancement of major areas. Hence also contributes to business benefit. Super predictor of IPL holds a good point in making journey of game more interesting.

FUTURE SCOPE:

Data Science is not only helpful for predicting the most favorable team for the tournaments but also helps generate valuable insights for other aspects.

Researchers have used Google trends to refer to data science for a deeper cricket match analysis. Certain Indian analytics companies like Cricket-21, play a huge role in data analysis for most global teams.

Data science helps us to extract knowledge or insights from data- either structured or unstructured- by using scientific methods like mathematical or statistical models. In the last two decades, it has been one of the most popular fields with the rise of all big data technologies. A lot of companies have been using recommendation

Sports is another field which is using data science extensively to improve strategies and predicting match outcomes. Cricket is a sport where machine learning has scope to dive into quite a large outfield. It can go a long way towards suggesting optimal strategies for a team to win a match or a franchise to bid a valuable player.

Opportunities in big data and sports analytics are now even greater than usual. The future of

machine learning is bright in the world of cricket. Big data has a vital role to play in decision-making for cricket, based on the available data. People don't chase cricket anymore. In fact, it is the sport that is running behind fans with data.

Machine learning technique WASP predicts final score

A machine learning technique called Winning and Scoring Prediction (WASP) predicts the final score in the first innings and estimates the chasing team's probability of winning in the second innings. And it works as a scoring predictor in the first innings of a match.

BIBLIOGRAPHY:

Kaggle site for downloading dataset
online resources

APPENDIX:

nil