Thistimedoneliao

by Ming Hun CHOW

Submission date: 05-Apr-2019 09:36AM (UTC+0800)
Submission ID: 1105736062
File name: Submissionpurpose_-_Copy.pdf (1.56M)
Word count: 9753
Character count: 50317

ANALYSIS OF NBA'S PARITY PROBLEM USING SUPERVISED LEARNING TECHNIQUE

By

CHOW MING HUN

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Science (Hons.) Applied Mathematics With Computing

Lee Kong Chian Faculty of Faculty of Engineering and Science

Universiti Tunku Abdul Rahman

April 2019

DECLARATION OF ORIGINALITY

I hereby declare that this project report entitled "**ANALYSIS OF NBA'S PARITY PROBLEM USING SUPERVISED LEARNING TECHNIQUE**" is my own work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : _____

Name : _____

ID No. : _____

Date : _____

APPROVAL OF SUBMISSION

I certify that this project report entitled "**ANALYSIS OF NBA'S PARITY PROBLEM USING SUPERVISED LEARNING TECHNIQUE**" was prepared by **CHOW MING HUN** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science (Hons.) Applied Mathematics With Computing at Universiti Tunku Abdul Rahman.

Approved by,

Signature : _____

Supervisor : _____

Date : _____

ACKNOWLEDGEMENT

It is an incredible opportunity given by University Tunku Abdul Rahman (UTAR) for me to finish this research. I had experienced a wide range of journals and sites to get more data to perfect this study.

I recognize with appreciation to Ms Choo Ley Ya, who has dependably been earnest and helpful in influencing me to comprehend the direction to finish this project and applied issues in my paper. Subsequently, I finished this project under the supervision and guidance of her.

Next, I am thankful to my family since they give me chance to pick UTAR to finish my tertiary education. In addition, I am also grateful to our companions and every single individuals who contributed their help in the fulfillment of my research project.

CHOW MING HUN

ANALYSIS OF NBA'S PARITY
PROBLEM USING SUPERVISED
LEARNING TECHNIQUE

CHOW MING HUN

ABSTRACT

Since the change in NBA playoffs format which split the league into East and West, there occurs a lack of parity between the two conferences. NBA's competitive balance has become a controversial topic. Nowadays, the West Conference is more competitive conference where the average quality of teams is clearly exceeding the East.

In order to analysis the issue, this study was using supervised learning technique, Naïve Bayes to predict the result which the model. Features selection was conducted for improving the model accuracy.

With suitable model applied, this project presented the prediction of the matches by the Round Robin Schedule's table. This analysis will be carried on by showing the winning rate for each team with the current playoff format and thus compare with the suggested format for the playoff, where the current playoff format is the elimination with best of 7 series for each round.

The evidences from the analysis will show the effect of the parity problem in NBA which to test the seriousness about the issue and hoping to create the awareness for all the members from NBA.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS / ABBREVIATIONS

32

NBA National Basketball Association

BAA Basketball Association of America

TOR Toronto Raptors

BOS Boston Celtics

PHI Philadelphia 76ers

CLE Cleveland Cavaliers

IND Indiana Pacers

MIA Miami Heat

MIL Milwaukee Bucks

WAS Washington Wizards

HOU Houston Rockets

GOS Golden State Warriors

POT Portland Trail Blazers

OKC Oklahoma City Thunder

UTA Utah Jazz

NEO New Orleans Pelicans

SAA San Antonio Spurs

MIN Minnesota Timberwolves

CHAPTER 1: INTRODUCTION

1-1 Background

In 1949, The National Basketball Association (NBA) was established by consolidation between the Basketball Association of America (BAA) and National Basketball League. However, NBA identifies only three seasons from BAA league as its history. These is because, the three seasons of BAA champion were competed in a best-of-seven arrangement. The 1947 and 1948 BAA playoffs were generally nearly match current NBA playoffs format. At the early year as 1947 and 1948, the playoffs format follow the best-of-seven elimination by separated into two conferences, Eastern and Western. Those achieved the champion in their conference will come to the last title match for the season.

September 2015, NBA reported changes of format for the NBA Playoffs 2016. The top eight groups in each conference were positioned all together by win-loss records to fit the bill. The criteria for playoffs' seeding and advantage of home had slightly improve, the situation of direct confrontation between the tied teams and the team who won the division championship. Series are played in the format that the home team will hosts games at first, second, fifth, and seventh rounds, while the rest will hosted by opponent team. Since 2014, this format has been used, but later after NBA decided to change from a 2–3–2 format which means the home team will conduct the game at their territory at first two and the last two rounds on October 23, 2013 due to the vote.

These playoffs' seeding are utilized to construct the bracket that decides the schedule for the league. Once it begin, section will be settled; hence no rearrangement for the consecutive matches. This is clearly different from the league of National Football League (NFL) and Major League Soccer (MLS) where the seeding proceeds throughout the tournament to ensure the strongest teams face the weakest one.

1-2 Problem Statement

Since the change in NBA playoffs format which split the league into East and West, there occurs a lack of parity between the two conferences. NBA's competitive balance has become a controversial topic. Nowadays, the West Conference is more competitive conference where the average quality of teams is clearly exceeding the East. These make the potential teams faced each other before reaching the final stage. For instance, the Thunder, one of the promising teams to reach the top, but met their rival, Utah in first round of playoffs, making them disqualified earlier before reaching the title match on season 2017/18.

Another problem arising from current format is the better team might be dropped out from joining the playoffs. This statement can be clearly supported on season 2017/18, when nine teams in the West and seven teams in the East achieved the top 16 before playoffs. However the Heat (from East) manage to enter the playoffs and Clippers (from West) was eliminated by the current format which spliting the league into two conferences.

1-3 Objective

The objectives of this project are:

1. Discuss the issues arising from NBA's parity problem.
2. Predict NBA's matches results using selected supervised learning technique.
3. Analyze NBA's matches' results with different playoff formats

1-4 Project Scope

In this project, comparison on the selected playoff formats will be discussed. NBA's parity problem arose from the current playoffs format will be investigated. Selected supervised learning technique will be used for predicting the matches' results to support this analysis.

CHAPTER 2: LITERATURE REVIEW

2-1 Parity Problem in NBA

Since 2014/15 NBA season until current season, the franchises who reached the NBA playoff's final were only the Golden State Warriors (Western Conference) and Cleveland Cavaliers (Eastern Conference). The arrangement is an epic showdown between a dauntless power and a stunning accumulation of ability, and it was making the playoff's final become more than the grudge match. There is because Lebron James wants to be the best player in the league and The Warriors wants to be the greatest team ever with the streak of wining.

It was on 2017, Dan (2017) stated that the current issue occur in NBA games is actually the parity. Due to the parity issue, NBA season leads to the very consistent seasons with non-aggressive playoff competition and fears of a disproportionate championship series. An unexpected increment in the salary cap allowed the Warriors to sign one of the most valuable players, Kevin Durant where such increments lead the parity issue become more serious. Apart than that, the draft lottery in NBA is aimed at aiding the low rank teams, but the unintended consequences are actually more severe. The bottom teams remains as the feeder because of its poor management, while the average teams showed the possibilities.

Thru the showdown between two strong houses, Aschburner (2018) showing that the two teams paying for the on-court success of their team which making the teams ranked as first and second in team payroll. The willingness of paying dollar to achieve the top team is clearly larger than the talent and strategies. By paying for the luxury tax which for the team who got the salary exceed the soft cap, it make the team able to recruit the more capable players for conquering the top by paying them as high as possible salary.

Kitano (2018) mentioned the playoff seeding system was making the balance off which also the parity issue. Other than the individual players' desire, the current playoff seeding system making the top 16 team to join for the NBA playoff by separated it to 2 conferences which contained the top 8 in each conferences. However, these selected top 16 teams were actually not always the real top 16 of rank in the league. From 2017/18 season of NBA playoff, the actual top 16 were actually 9 teams from West and 7 teams from East,

causing the matchups losing some of the possibilities. Kitano's word also supported by NBA commissioner, Adam Silver's comment during NBA All-Star Weekend 2018, he mentioned, "You also would like to have a format where your two best teams are ultimately going to meet in the Finals. You could have a situation where the top two teams in the league are meeting in the conference finals or somewhere else. So we're going to continue to look at that. It's still my hope that we're going to figure out ways."

2-2 Supervised Machine Learning

According to Taiwo (2010), machine learning is tied in with structuring algorithms that enable machine to learn. The learning process not really includes consciousness. However learning involves finding measurable regularities or different examples in the information. He also mention about learning technique of the machine will take after the human's way to study a machine learning assignment. Thus, the methods of the machine learning give the understanding towards the trouble of learning process in various conditions.

Machine learning classification requires from settings of the parameters and the number of cases for the data set. It's not only the time to be care in building the model of the algorithm but also the precision and the correct classification are considered too. Thus, the best learning algorithm for a specific data collection does not ensure the accuracy and precision for another set of data which the characteristics are sensibly not quite the same as the other. However the researchers also mention about, the key inquiry when managing machine learning algorithms isn't whether which algorithm is superior towards others, yet to find out under some certain conditions , there is some specific method can fundamentally outflank other methods on given that some real life application issue. Meta-learning introduced since it utilizes a lot of features, called meta-attributes which the attribute represent to the specification of learning jobs, and analysis on the relationships between the specification and its performance. (F.Y et al., 2017)

Muhammad and Yan (2015) stated that supervised machine learning is widely used in various areas. Due to the nature of machine learning, each algorithm of machine learning consist their own characteristics which are the benefits or weaknesses. So, types of the task

nature are the key to choose the algorithm in machine learning. On this study, support-vector Machine and Neural Networks had the slightly overwhelming result when settling with multi dimensions and continuous attributes of the datasets. Logic-based type of algorithm eventually work superior when work with features with the absolute features. For Neural Network classifier, a huge datasets was needed for training in order to achieve the optimistic accuracy of the model; however Naive Bayes only require small amount of value for the training data. Lastly, they conclude that the bigger size data did bring negative impact and it will lead wipe options for machine learning classifier and particularly to the deep learning method. So, it had become a mainstream technology for different type of application target for past few years.

This paper summarizes the fundamental aspects of couple of supervised methods. The main goal and contribution of this paper is present the overall result of machine learning and provide machine learning techniques. (Nasteski, 2017)

Hashemi and Karimi (2018) discussed about the inequality in training dataset will occur in various tasks because some training data are measured by different perspective or machine's characteristics. For non-weighted machine learning methods, they are intended for similarly critical preparing tests: first, the expense of misclassification is equivalent for preparing tests in parametric characterization systems; second, residuals are similarly imperative in parametric relapse models and when casting a ballot in nonparametric classifier and regression method, preparing tests with the level of weights or their weights are resolved inside by kernels in the element space, in this manner no outside weights. In this work, they developed the weighted versions of Bayesian predictor, perceptron, multilayer perceptron, SVM, and decision tree and compare the results with their non-weighted versions.

The objective of this instructional exercise was presenting key models, algorithms, and inquiries identified with utilization of enhancement strategies to investigating the issues emerging towards learning technique. They start by determining a detailing of a supervised learning issue and show how it prompts different utilization issues, contingent upon the specific situation and hidden suspicions. They also examine a portion from particular highlights of enhancement issues, concentrating towards instances of logistic regression and train progress for deep neural systems. Last part of practical concentrated around improving for the algorithm, first for convex logistic regression. At last, they talk about their methodologies able to utilize to the neural system’s training, underscoring challenges which

will emerge from the complexity, non-convex pattern of models. (E. Curtis and Scheinberg, 2017)

From overview and examination on correlation between the machine learning classification algorithms which are Decision tree, Bayesian, K-Nearest Neighbor. It demonstrates that the Decision Tree's methods were progressively exact and consists slightly low blunder rating and also much simpler calculations when contrasted with the other two methods. The after effect when utilized in WEKA towards the equivalent target sets demonstrated that the decision tree beats the other two models. Bayesian classification having indistinguishable exactness from the decision tree yet the K-Nearest Neighbor didn't contribute great outcomes. The examination has appeared every algorithm contains their very unique advantages and disadvantages just for its very self-region of usage. Neither one of methods was able to fulfil all of the tasks. Contingent upon necessities, explicit algorithm will be picked. (Jadhav and Channe, 2016)

There are few real life cases which also applied the machine learning technique in order to help their analysis the issues. For this article, the main target of this article which investigate by Kannadasan, Prabakaran and Chandraa (2018), was to determine and predict the airline delays caused by different elements. The purpose for their research was to decrease the impact of flight delays which will affect fellow parties, mainly economical for commuters, airline industries and airport authorities. Besides that, in the domain of sustainability, the increased used in fuel consumption and gas emissions will also cause the environment issues. To carry out the predictive analysis, they were using the Regression Analysis for giving a detailed analysis of the result of individual airlines, airports, and for evaluated choice. Additionally, aside from the appraisal identified with the travellers, this investigation will likewise help in decision making procedures methods necessary for each significant player in the air transportation system.

In this article, the researchers had stated this problem of software risk identification and assessment by utilizing of two techniques of supervised learning methods such as neural networks with back propagation and logistic regression with regularization. The proposed strategy consists ton of guarantee to submit the errands of risk controlled simpler as well as increasingly accurate. The proposed methods are simple, deployment friendly, as well as efficiency in the evaluations of the experiment. For continuous work, they are investigating

different expansions towards essential positioning instrument in order to give better risk assessment and make the framework progressively strong. (J. et al., 2014)

By the study of Bhavsar and Ganatra (2012), the examination of the degree notable classification algorithms had conducted for the analysis. The point to carry the investigation was to get familiar with major thoughts as well as locate recent examine issues, thus it can provide the help to different specialists just for understudies who are completing a propelled seminar on machine learning classifier. This current study had appeared every algorithm has its own advantages and disadvantages and burdens just as its own zone of execution. Neither one of methods was able to fulfil all of the tasks. The way to research on the machine learning classifier, can be worked by a combination with at least 2 classifiers by joining the quality.

In this study, Fraussen, Graham and Halpin (2018) acquainted a novel hypothetical methodology to address this vital inquiry, in particular prominence. They contend that, in the administrative field, prominence can be operationalized as gatherings being referenced deliberately utilized as an asset by chose authorities as they debate policy matters. Besides that, they apply a machine learning answer for dependable survey which bunches are conspicuous among officials. They show this novel strategy depending on a dataset of mentions of 1300 national interest groups in parliamentary debates in Australia over a six-year time frame (2010-2016).

Dulhare (2018) inferred that the proposed model is powerful and effective to improve the precision of the Naive Bayes classifier utilizing the particle swarm optimization for feature subset determination which accomplishes comparable or shockingly better characterization execution. The objective was effectively accomplished by building up a novel calculation amplifying the order execution and limiting the quantity of features. From recreation results, it investigated that this algorithm could naturally develop a component subset choice with a less number of features and increment classification performance than utilizing every one of the features of a dataset. Later on the future, he wanted to create proposal framework for early prediction of coronary illness finding. Likewise the utilization of particle swarm improvement for features determination on datasets with countless can likewise be concentrated to take note of the different parts of particle swarm advancement in feature choice.

Chapter 2 Literature Review

The scientists of this article made utilization of a Bayesian regulated learning approach in evaluating American options by means of Monte Carlo simulations. They first present Gaussian procedure regression approach for American alternatives pricing and analyze its execution in assessing the continuation value with the Longstaff and Schwartz algorithm. Furthermore, they investigate the control variants system in blend with Kriging to additionally enhance the estimation of the continuation value. They referenced that this technique allows to reduce drastically the standard errors and to enhance the solidness of the Kriging approach. They utilized American put options on a stock whose elements is given by Heston model, and utilize European options on same stock from control variates for illustrative purposes. (Mu et al., 2018)

This study described how gossip and rumours can be characterized as a flowing unconfirmed story or a dicey truth. Gossip initiators look for social networks defenceless against illimitable spread, accordingly, online media turns into their stage. Consequently, this deception forces huge harm to people, associations, and the legislature, and so on. Existing work, dissecting fleeting and semantic attributes of bits of gossip appears to give abundant time for rumour propagation. In the interim, the tremendous upheaval of information via social media, considering these attributes for each tweet turns out to be spatially complex. Along these lines, in this article, a two-fold supervised machine-learning system is suggested that recognizes bits of rumours by separating and afterward breaking down their etymological properties. This technique endeavours to automate filtering via preparing various grouping algorithms with precision higher than 81.079%. At long last, textual characteristics on the filtered data had been applied and rumours are recognised. The adequacy of the proposed structure is appeared broad examinations on more than 10,000 tweets. (Thakur et al., 2018)

Because of the development of Internet get to, the requirement for secure and dependable systems has turned out to be increasingly critically. The modernity of system assaults, just as their seriousness, has likewise expanded as of late. All things considered, an ever increasing number of associations are getting to be defenceless against assault. The point of this study was to managing network assaults using neural systems, which can provide better detection rating and insignificant false alert with only short period of time needed. Current study focused around two classification types which a solitary class, and a multi class, where the classification of assault is also distinguished by the neural network. Extensive examination is conducted to survey the interpretation of representative information,

dividing of the training data and the multifaceted the architecture. (Fares, Sharawy and Zayed, 2011)

Other than analysis the cases, there was the study about improving the performance of the model. In this paper, Frank and Bouckaert had identified a potential deficiency of multinomial naive bayes in the context of unbalanced datasets and shown that per-class word vector normalization presents a way to address the problem. Their empirical results showed that normalization can indeed significantly improve performance. They had also shown that MNB with class vector normalization is very closely related to the standard centroid classifier for text classification if the class vectors are normalized to unit length, and verified the relationship empirically. (Frank and Bouckaert, n.d.)

Another study about improving performance of the model which utilizes spatial filtering of event information with the point of diminishing overfitting to examining bias in cological niche models (ENMs). Inspecting bias in geographic space prompts territories that may likewise be one-sided in natural space. As a primer test tending to this issue, they utilized Maxent, bioclimatic factors, and event areas of an extensively dispersed Malagasy tenrec, Microgale cowani (Tenrecidae: Oryzorictinae). They demonstrated the abiotically appropriate territory of this species utilizing three particular datasets: unfiltered, spatially separated, and tenuous unfiltered areas. They also determined assessment area under the curve (AUC) by utilizing the contrast among alignment and evaluation AUC, and exclusion rates to evaluate overfitting and model performance. Models made with the sifted dataset demonstrated lower overfitting and preferred execution over the other two suites of models, having lower exclusion rates and evaluation AUC of differences, and a higher AUC for evaluation. Also, the tenuous unfiltered dataset performed superior to the unfiltered one for three assessment measurements, likely in light of the fact that the bigger one fortified the biases. These outcomes show that spatial separating of event territories may enable biogeographers to create better models. (Boria et al., 2014)

2-3 Supervised Machine Learning on Sports

At 2006, Joseph, Fenton and Neil constructed a test of performance between the different types of machine learning technique. They used Bayesian Networks to check the performance by compared with other machine learning techniques for predicting the outcome of matches played by Tottenham Hotspur Football Club at the period under of 1995–1997. The extra techniques used were: MC4 (a decision tree learner); Naïve Bayesian learner; Data Driven Bayesian and a K-nearest neighbour learner. Their results explained that Bayesian Networks got the better performance compare to the other techniques for predicting accuracy. Other than predicting accuracy, Bayesian Networks also having the ability to predict the data with limited learning data. Moreover, Bayesian Networks was simple to build and repeatable on similar problems. (Joseph, Fenton and Neil, 2006)

Another approach to predict the result which Kvam and Sokol (2006) have proposed a logistic regression and Markov chain model. By their method, the teams pretended as a single state of the Markov chain and if there is a team better than another, it will present the state transitions. To estimate transition probability parameters from the data, they used logistic regression to fulfil this task.

After 4 years period, Brown and Sokol (2010) have presented an updated version using Bayesian estimates. By replacing logistic regression and Markov chain model with empirical Bayes and ordinary least squares, the after results show that changing the logistic regression with the two empirical Bayes models obtain a slightly improvement when the probabilities are jointly conditioned.

Later, there was also some model for predicting the result by different type of model. One of the significant researches had been done by Loeffelholz, Bednar and Bauer (2009). They used the neural network to predict the result of NBA. Total of 620 NBA games data were collected for the training of the model such as feed-forward, radial basis, probabilistic and generalized regression neural networks. Results obtained by them were able to predict accurately with 74.33% of the average. (Loeffelholz, Bednar and Bauer, 2009)

Jasper Lin, Logan Short and Vishnu Sundaresan in "Predicting National Basketball Association Winners," were achieved the range of accuracies which 63.3 to 65.1% with

logistic regression, adaptive boost, random forest and support vector machines. Box score statistics was used from games starting with the 1991-1992 season through the 1997-1998 season. Their conclusion is that the past results had played the important role in predicting coming results. However their accuracy dropped, when the win records exclude from their training. Thus the box score statistics get no identical evidence to increase the accuracy of predicting for the winning. (Jasper, Logan and Vishnu, 2014)

"Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics" by YuanHao, analyzes the correlation between individual player's statistics and their team's performance, and develops a prediction model that can be used to forecast regular season results of NBA teams based on common player statistics. (YuanHao, 2015)

CHAPTER 3: METHODOLOGY

3-1 Domain Understanding

At the beginning of stage, the study about this sport, basketball was done to understand the rules and the condition of wining this sport. The factors are potentially affect the result of the games also be investigate for further predictions.

3-2 Data Preparation

The consideration for selecting data was important for this analysis. Previous studies generally using team level data for their training set. Thus this analysis will include individual player data, which contains the statistics on each single player. Player level data will separate individually and joined with the match data to ensure every match has the individual statistics as attributes in the data set. Hence, it gained the merits to study the impotency of players' actions or presence in terms of teams.

The data is obtained from Basketball-Reference which is the database consists of all the statistical results in major basketball game. There are 24692 players' statistics from the collections with 51 attributes for each player. Selection of the players to forms the team will be executing for further testing.

Total 3 years games which are 2016, 2017 and 2018 season games used for training model. Each year's data consists of 1230 games. Both teams in each games' data was insert with 13 players' data in each team since in every matches, NBA's rules only allow 13 players involved. In the player' data, some attributes was selected according to the k highest scores which are minute play, position and field goal rate. Apart then the rules, team performance usually depends on first 10 players but the other 3 included to ensure model accuracy, there are some teams (Lakers, 76ers and etc.) actually got the strategic for higher frequency of rotating players during the game, it is to help the players stay at the peak performance. Other

might use the last 3 members slot as newbie training in the real game. Thus, the minute of played was taken as important attribute for evaluating the result.

Table 3.1: Sample Statistic of Player

TeamWin	TeamSide1	HR	MP_PG_H1	FG_PG_H1	RAT_PG_H1	POS_PG_H1
0	1	65.9	28.5	43.9	78	0

Where,

TeamWin = Home win: 1; Away win: 0

TeamSide1 = Home: 1; Away: 0

HR = Home Rating; AR = Away Rating

MP = Minute Played

FG = Field Goal (100%)

RAT = Player Efficiency Rating

POS = Position Played of the player

Position: 0 as PG/ Point Guard

1 as SG/ Shooting Guard

2 as SF/ Small Forward

3 as PF/ Power Forward

4 as C/ Center

Column TeamWin is the result for every matches, it is use to check the predictive result and actual result for the accuracy. Total 2 teams in each row of data, thus TeamSide1 and TeamSide2 use for determine the team which are home or away.

In the table above (Table 3.1), only the first player's data was included, but 13 players included too in the actual dataset, which H1 meaning of home first team's players (5 members), H2 as second team (5 members) and the other extra player E1,E2 and E3. The away team contain the exact same characteristic as home team, only the first team's and second team's player representative as A1 and A2. Extra team member consists the extra attribute call POS which is the position of the player.

3-3 Naive Bayes Classifier

For supervised learning technique, Naive Bayes is an individual from basic probabilistic classifiers which using Bayes' theorem with solid independence supposition between the characteristics.

3-3-1 Naïve Bayes

Naive Bayes is a straight forward technique for building the classifiers. According to the common theory, Naïve Bayes classifiers expect that the value of some attribute is independent with any other attribute together with the class variable. For instance, a bloom might be viewed as a sunflower on the off chance that it is yellow, 30cm length, and around 15 cm in distance across. A Naive Bayes classifier considers every one of these features to contribute identically to the likelihood that this blossom is a sunflower, whether there had any connections between the colour, length, as well as diameter features.

Due to the certain kinds of probabilistic models, Naive Bayes classifiers are working effectively inside the supervised learning environment. In several reasonable applications, estimation of the parameter for Naive Bayes models used the strategy for maximum

probability. Due to this scenario, anyone can construct the Naïve Bayes models without letting any Bayesian probability involved to the models, as well as the methods of Bayesian.

Even with the naïve plan and evidently oversimplified supposition, Naive Bayes classifiers are still working perfectly in various types of complex situations that occur in the real word. An examination of the issue about the Bayesian classification demonstrated that occur sound hypothetical purposes for the clearly impossible viability of Naive Bayes. (ZHANG, 2005)

An advantage to choose Naive Bayes is due to the requirement of the training data is actually less significant. On the off chance that the independence assumption holds, it will converge faster than discriminative models such as logistic regression.

Naive Bayes is a conditional probability model. Given an issue example to be characterized, represented by a vector $x = (x_1, \dots, x_n)$ representing some n independent factors, it appoints to this occasion probabilities

$$p(C_k | x_1, \dots, x_n)$$

For each of K possible outcomes or classes C_k. (Murty and Devi, 2011)

The issue with the above definition is that if the quantity of features n is substantial or if that an element can take on an expansive number of qualities, at that point putting together such a model with respect to likelihood tables is infeasible. Subsequently reformulate the model to make it progressively tractable. The conditional likelihood can be disintegrated by using Bayes' theorem as

$$P(C_k | x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

It can be explained by plain English, using Bayesian probability terminology

$$posterior = \frac{prior * likelihood}{evidence}$$

By and by, there is interest just in the numerator of that division, due to the fact that the denominator does not rely upon C and the estimations of the features x_i are given, with the goal that the denominator is effectively consistent. The numerator is proportional to the joint likelihood model

$$p(C_k, x_1, \dots, x_n)$$

By utilizing chain rule for the replication of the definition of conditional likelihood, it can be rewritten as follows:

$$p(C_k, x_1, \dots, x_n) = p(x_1, \dots, x_n, C_k)$$

$$= p(x_1|x_2, \dots, x_n, C_k)\, p(x_2, \dots, x_n, C_k)$$

$$= p(x_1|x_2, \dots, x_n, C_k)\, p(x_2|x_3, \dots, x_n, C_k)\, p(x_3, \dots, x_n, C_k)$$

$$= \dots$$

$$= p(x_1|x_2, \dots, x_n, C_k)p(x_2|x_3, \dots, x_n, C_k) \dots p(x_{n-1}|x_n, C_k)p(x_n|C_k)p(C_k$$

Assume that each feature x_i is conditionally independent of every other feature x_j for $j \neq i$, given the category C_k. This means that

$$p(x_i|x_{i+1}, \dots, x_n, C_k) = p(x_i|\, C_k$$

Thus, the joint model can be explained as

$$p(C_k|\, x_1, \dots, x_n) \propto p(C_k, x_1, \dots, x_n)$$

$$= p(C_k)\, p(x_1|\, C_k)\, p(x_2|\, C_k)p(x_3|\, C_k) \dots$$

$$= p(C_k)\, \prod_{i=1}^{n} p(x_i|\, C_k)$$

Where $\propto$ denotes proportionality.

This means that under the above independence assumptions, the conditional distribution over the class variable C is:

$$p(C_k|\, x_1, \dots, x_n) = \frac{1}{Z}\, p(C_k) \prod_{i=1}^{n} p(x_i|\, C_k)$$

$$Z = p(x) = \sum_k p(C_k)\, p(x \mid C_k)$$

Where the evidence Z is a scaling factor dependent only on $x_1, \dots, x_n$, that is, a constant if the values of the feature variables are known.

3-3-2 Gaussian Naive Bayes

While managing with the continuous information, the classic supposition which assume the constant qualities related to each individual characteristic, are dispersed by Gaussian distribution. Assume that data for training consists of continuous features, x. Firstly, classified the training datasets, and then figure the mean and variance of x in each class. Let μ_k be the mean of the values in x related with class C_k, and let σ_k^2 be the variance of the values in x related with class C_k. Suppose there is observation value v. Thus, the likelihood distribution of v given a class C_k, $p(x = v|C_k)$, can be expressed by inputting v into the equation for a Normal distribution parameterized by μ_k and σ_k^2. That is,

$$p(x = v|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

3-4 Assumption

According to Naïve Bayes theorem, the features, $x_1, x_2, \ldots, x_p$ where p is the total number of the features, are independent. Thus, the independent attribute of the player selected from their statistic data. Due to the natural of the Naïve Bayes, each of the variables took as same weight during the analysis, so assume that the players' performance are independent depends on their position, with 5 of the starter, 5 from main bench team and other 3 as the bench's substitute, and each players contribute identical impact towards the result. Under this assumption, the model processed to get the result.

3-5 Evaluating Model

3-5-1 Model deployment

Ideally, each round of games will be analysis by the technique. The comparison between the model and the actual result will be done to achieve the accuracy of the model. After ton of instances for the constant average accuracy, the parity problem will be analysis between after and before the suggested method applied on the parity

3-5-2 Measuring Performance

The match results will be classified into two category of wining state: home and away. Simple classification matrix will be deployed to identify the accuracy.

By using Gaussian Naïve Bayes, the accuracy was able to reach 70.8% with only small amount of time which both training and prediction time taken less than one second.

Table 3.2: Confusion Matrix of the Model

		True Class	
		Away	**Home**
Predicted Class	**Away**	340	180
	Home	178	532

13

Table 3.3: Classification Report

	precision	recall	f1-score	support
0	0.65	0.66	0.66	518
1	0.75	0.75	0.75	712
micro avg	0.71	0.71	0.71	1230
macro avg	0.7	0.7	0.7	1230
weighted avg	0.71	0.71	0.71	1230

3-6 Preliminary Analysis for NBA Matches

After the modeling, the prediction for the matches will be presented by the Round Robin Schedule's table. The accuracies of the model compare to the real result will calculate for the analysis purpose.

With the suitable accuracy, the analysis will be carried on by showing the winning rate for each team with the current playoff format and thus compare with the suggested format for the playoff, where the current playoff format is the elimination with best of 7 series for each round.

The evidences from the analysis will show the effect of the parity problem in NBA which to test the seriousness about the issue and hoping to create the awareness for all the members from NBA.

Suggested Round Robin Schedule's table will be generated, as the figure:

Team Name		1	2	3	4	5	6	7	8	Points	Rank
1	Houston		1	1	1	1	1	0	1	6	2
2	Minnesota	0		1	0	0	1	0	1	3	4
3	Oklahoma City	0	0		0	0	0	0	0	0	8
4	Utah	0	1	1		1	1	0	1	5	3
5	Portland	0	1	1	0		0	0	1	3	4
6	New Orleans	0	0	1	0	1		0	1	3	4
7	Golden State	1	1	1	1	1	1		1	7	1
8	San Antonio	0	0	1	0	0	0	0		1	7

Figure 3.1 Sample NBA Playoffs schedule by round robin system

The winning teams will get 1 and lose for 0, thus the ranking will depends on the total points earning after the suggested format finished. Thus, the team has the best potential will remain at the end for the champion.

Figure 3.2 NBA Playoffs 2018 Brackets

The current NBA Playoffs format which is the best-of-7 elimination matches between Western conference and Eastern Conference. The top of the both conferences will compete at the final of playoffs to achieve the champion title for the current season.

CHAPTER 4: RESULT AND DISCUSSION

4-1 Analysis of NBA Playoff Result

4-1-1 Playoff Qualifying on 2017-18 NBA Playoffs

Before go into the main analysis of the different method for playoffs, there was the quick summary for the playoff qualification.

Toronto turned into the principal group which secure the position to join the playoff on March of 7. In the month which the 30th day, Houston secured the West's champion by finishing the streaks of three consecutive years of Golden State Warriors. Houston also secured the top result in NBA history on the next day after it beats the Golden State record. Minnesota vanquished Denver with the score 112 and 106 respectively in extra time to secure the last chance to join the playoff in the west conferences. This likewise finished 13-year's curse for Minnesota which is never joined the playoff since 2003. Los Angeles missed the chance to join the postseason because lost to Denver on their last game. New York, Los Angeles, or Chicago never made it in to the playoff ever since 1960 which at least one of them will join the playoff.

Chapter 4 Result and Discussion

Seed	Team	Record	Clinched			
			Playoff berth	Division title	Best record in conference	Best record in NBA
1	Toronto Raptors	59–23	March 7	April 6	April 6	—
2	Boston Celtics	55–27	March 8	—	—	—
3	Philadelphia 76ers	52–30	March 26	—	—	—
4	Cleveland Cavaliers	50–32	March 22	April 10	—	—
5	Indiana Pacers	48–34	March 25	—	—	—
6	Miami Heat	44–38	April 3	April 11	—	—
7	Milwaukee Bucks	44–38	April 4	—	—	—
8	Washington Wizards	43–39	March 31	—	—	—

Figure 4.1: Summary of Eastern Conference

Seed	Team	Record	Clinched			
			Playoff berth	Division title	Best record in conference	Best record in NBA
1	Houston Rockets	65–17	March 11	March 15	March 29	March 29
2	Golden State Warriors	58–24	March 12	March 15	—	—
3	Portland Trail Blazers	49–33	April 1	April 11	—	—
4	Oklahoma City Thunder	48–34	April 9	—	—	—
5	Utah Jazz	48–34	April 8	—	—	—
6	New Orleans Pelicans	48–34	April 10	—	—	—
7	San Antonio Spurs	47–35	April 10	—	—	—
8	Minnesota Timberwolves	47–35	April 11	—	—	—

Figure 4.2: Summary of Western Conference

The bolded team win at the current round. The numerical located on left of the team name show the seeding position in their conference where E means east and W mean west. The right hand side of the team's name show the total number of game that win in the current round. The champions of each type of division are highlighted by an asterisk. The teams with the italics style are the team having the home court benefit and also the better seeding team.

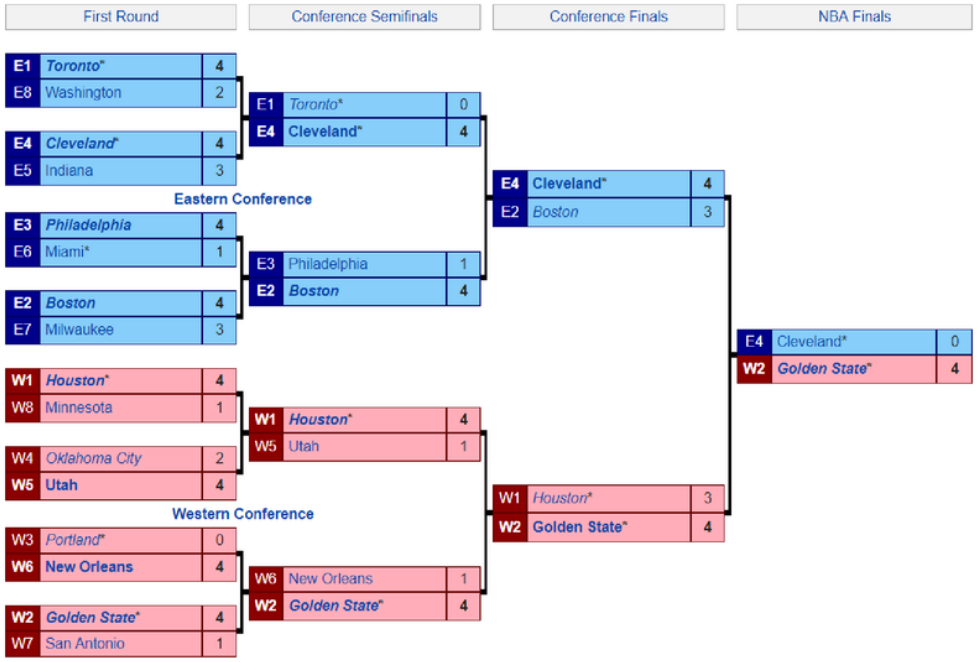

Figure 4.3 Playoff Bracket NBA Season 2017-18

4-1-2 Best-of-seven Elimination

The playoff of the NBA is actually the postseason championship. The playoffs started on April 14 and ended on June 8 at the conclusion of the 2018 NBA Finals.

For each conference, top eight teams will be selected for the qualification of the current playoff. The seedings are based on the record of each team which is the total win by ranking.

Every conference's section is fixed, so there will be no reseeding for the playoff. All matches are conducted by best of seven elimination arrangement which the team need to obtain four successes in order to advance to the following round. This arrangement is effective to all the rounds in playoff which are in a 2 games at team A home court, 2 for team B, the rest matches will conduct in each team's home one more time and the last will get by the team with home court advantage. Home court advantage will be given for the group with the best season record, rather than to those top seeding group. On the off chance that two groups with a similar result face each other in the match, the standard tiebreaker rules will be utilized. The standard for deciding home court advantage in the NBA Finals is winning rate, at that point straight on record, trailed by record versus opposite conference.

4-1-3 Round-robin Tournament

A round-robin competition which likewise called all-play-all competition is a challenge in which every contender meets every other competitor for the matches. A round-robin stands out from an elimination competition, in which members are disqualified after a specific number of losses.

In a single round-robin plan, every member plays each other member only once. If that every member plays all others twice, this is regularly called a double round-robin. Double round-robin used when all members play each other more than twice, and is never utilized when one member plays others an unequal number of times.

In games with countless matches per season, double round-robins are normal being used in the tournament. In the current era, most of the football leagues are sorted out on a double round-robin premise, in which each group plays all others in its association once at home and once away. This framework is also utilized in capability for major competitions, for example, the FIFA World Cup and the continental competitions. Apart than the football league, there are also round-robin bridges, chess, Go and Scrabble competitions which also utilized this rule system. The World Chess Championship chose in 2005 and in 2007 on an eight-player double round-robin competition where every player faces each other player once as white and once as dark.

The champion, in a round-robin competition, is the candidate that successes the most achieve. In the hover of death, it is conceivable that no champion rises up out of a round-robin competition, regardless of whether there is no draw.

Generally, round-robin rule based can be considered as the fairest method to decide the top from a group of the contesters. Every challenger, regardless of whether player or group, has meet possibilities against every single other opposite teams on the grounds that there is no earlier seeding of candidates that will block a match between some random pair. The luck factor was seen not so effective when contrasted with a knockout system since a couple of poor performances need not disable the probability of all the contesters to reach the champion.

The records of members are performing a more accurate result since they represent to the outcomes over a longer time towards a similar resistance. Apart than study on the best team, this method can used for determine the poorest teams. This is also handy to decide the last position for all the members, from most capable to the weakest, for reason to qualify the abilities for another level or tournament, and also the cash rewards. Under round-robin rule based, the champion is generally showed as the top of the best team in the tournament, rather than of the elimination rule's victors.

The scheduling algorithm for round-robin system can be expressed by, let n be the volume of candidates, pure round-robin competition will need $\frac{n}{2}(n-1)$ games to execute.

4-1-4 Application of Double Round-robins for NBA Playoff 2017-18

Double Round-robins was use to ensure the fairness of the court advantage which giving the chance for both team playing at own court once between the same team competing.

Result for winning team will gain 1 point and minus 1 for loss per game. The result table showed at below:

	TOR	BOS	PHI	CLE	IND	MIA	MIL	WAS	HOU	GOS	POT	OKC	UTA	NEO	SAA	MIN	HScore	Total	Rank
TOR		1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	13	23	3
BOS	0		1	1	0	1	1	1	0	0	0	0	1	1	1	0	8	15	8
PHI	0	0		1	0	1	1	1	0	0	0	0	1	0	1	0	6	10	11
CLE	0	0	0		0	1	0	0	0	0	0	0	0	0	1	0	2	2	15
IND	0	1	1	1		1	1	1	0	0	1	0	1	1	1	1	11	19	6
MIA	0	0	0	1	0		0	0	0	0	0	0	0	0	1	0	2	2	15
MIL	0	0	1	1	0	1		0	0	0	0	0	1	0	1	0	5	9	13
WAS	0	0	1	1	0	1	1		0	0	0	0	1	0	1	0	6	11	10
HOU	1	1	1	1	1	1	1	1		0	1	1	1	1	1	1	14	28	2
GOS	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	15	30	1
POT	0	1	1	1	1	1	1	1	0	0		0	1	1	1	1	11	19	6
OKC	1	1	1	1	1	1	1	1	0	0	1		1	1	1	1	13	23	3
UTA	0	0	0	1	0	1	0	0	0	0	0	0		0	1	0	3	5	14
NEO	0	0	0	1	0	1	0	0	0	0	0	0	1		1	0	4	10	11
SAA	1	1	1	1	1	1	1	1	0	0	1	1	1	1		1	13	13	9
MIN	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1		13	21	5
AScore	10	7	4	0	8	0	4	5	14	15	8	10	2	6	0	8		A = Away	H = Home

Figure 4.4 Round-robin Result for NBA Season 2017-18

According to the generated result, the ranking was shown as figure 4.5, both result show at the figure for easy comparison, elimination result categorize by boxes which the lower group eliminated at first round of playoff, second lower was the defeated team at the semi-final of the conferences and the top 3 boxes are the champion of the season, second place and the defeated at conferences final.

Rank	Team (Round-robin)	Team (Elimination)	Rank
1	Golden State Warriors	Golden State Warriors	1
2	Houston Rockets	Cleveland Cavaliers	2
3	Toronto Raptors	Houston Rockets	3
3	Oklahoma City Thunder	Boston Celtics	3
5	Minnesota Timberwolves	Toronto Raptors	4
6	Indiana Pacers	Philadelphia 76ers	4
6	Portland Trail Blazers	Utah Jazz	4
8	Boston Celtics	New Orleans Pelicans	4
9	San Antonio Spurs	Washington Wizards	5
10	Washington Wizards	Indiana Pacers	5
11	Philadelphia 76ers	Miami Heat	5
11	New Orleans Pelicans	Milwaukee Bucks	5
13	Milwaukee Bucks	Minnesota Timberwolves	5
14	Utah Jazz	Oklahoma City Thunder	5
15	Cleveland Cavaliers	Portland Trail Blazers	5
15	Miami Heat	San Antonio Spurs	5

From Western Conference	
From Eastern Conference	

Figure 4.5 Ranking of Round Robin Result

4-1-5 Comparison between Different Rule Based

After the analysis for the NBA playoff at season 2017-18, the result of the analysis had shown at the last session. According to the result of double round-robins rule, Golden State eventually stay at the first place which having the same result of the elimination rule. However, there is slightly different for the first fourth places.

From the actual result, the teams involved in the final game at each conference were Cleveland and Boston at the eastern conference, while Houston and Golden State had the last battle at the western conference. These can considered as the first fourth places under the elimination rule. Nevertheless, the matches under double round-robins rule produced the result for the first fourth were Golden State, Houston, Toronto and Oklahoma City; the last 2 teams gained the same score under the rule, so they were both at the third place. Based on the result, there was only Golden State stayed at the top four, and the other teams never joined to the final at their conferences, except Houston.

Oklahoma City was actually knocked out early at the first round of the elimination by Utah, the team actually having the outstanding performance and beat the Oklahoma City which got the 2^{nd} place on the analysis. Due to this situation, this potential team under the double round-robin rule actually defeated by their rivals. These rivals, Utah actually gained 14^{th}. In this situation, it showed that in the elimination rule might knock those potential teams at the early stage before the actual final. Toronto was also one of the potential team to reach the champion, but defeated by Cleveland at the semi-final in eastern conference, and Cleveland only ranked 15^{th} as the last team as well as Miami sharing the same rank.

Other than the difference at the top performance, Minnesota also reached higher ranking in the round-robin rule based which is the 5^{th} place, but it lost to 2^{nd} place, Houston at the first round of elimination. Philadelphia beat Miami at the first round of elimination game which is quite fair according to the analysis result which they were ranked as 11^{st} and 15^{th}; however Philadelphia suffered defeat by the 8^{th} place, Boston on the second round of the elimination.

There was another game in eastern conference, Milwaukee lost to Boston which Milwaukee lost the chance to continue competing at the playoff due to meeting the rival at the early stage. However, there were some team such as Utah and Cleveland at least managed

to get into the semi-final of the conference and these teams had the lower ranking compare to Milwaukee.

At the first round elimination in western conference, Utah actually having the outstanding performance and beat the Oklahoma City which got the 3rd place on the analysis. However, the outstanding performance of Utah never last longer, because it get beaten up by Houston at the semi-final. Back to the first round in West, Portland who ranked 6th also lose to New Orleans, 11th which made it stop at the first round of playoff. San Antonio who actually scores the 9th place, but it met the champion, Golden State at the first round, thus making San Antonio stay behind the playoff but pushed the Houston to the final at the conference.

In conclusion, there was slightly different visually of the result between the actual one, elimination based and the round-robin rule. The comparison result clearly showed that, the western conference's team are generally stronger than the East which proven by the round-robin rule base and there were 5 teams of West and 3 teams from East which reached the top 8 placed.

4-2 Issue of Modeling Result compare to Actual Result

After the analysis, there is the huge difference between the actual result and the analysis one, the ranking of the Cleveland. The main possible reason is because the efficient rating for the players in the team. Lebron James, who is the NBA all-star player and having the extra ordinary scoring than his other team member, made the average team performance slightly lower than other team. Due to the Naïve Bayes, it compare each of the player or each of the attribute 1 by 1, thus one of the outstanding result wasn't enough to affect the game result because the current NBA teams, at least having the quite balance team build than Cleveland.

Lebron James was conveying the Cavaliers in a notable manner. LeBron James' job as the Cavaliers' leading man is nothing unexpected. This issue never happen before, that the team in the playoffs has a player's supporting great amount of contribution on the team to carry the rest of the member. For this current season's playoffs, James has represented almost 33% of his group's box-score statistics: rebounds, steals, points, blocks and assists. Since 1974, These five statistics attribute have been kept reliably and no player ever since to contributed such higher share to his team's playoff result.

James 2018	Total	Percent of team total
Points	725	34%
Rebounds	193	22
Assists	190	47
Steals	30	24
Blocks	22	24

Figure 4.6 Lebron James's Statistical Data in Team

Clearly it can be seen, James has contributed 32 percent of the Cavaliers' statistical data. However, the Warriors' statistical leader was Kevin Durant, at only 22 percent which also due to the Golden State's team consisted much All-Stars players who actually can share the contribution towards the team's statistical data.

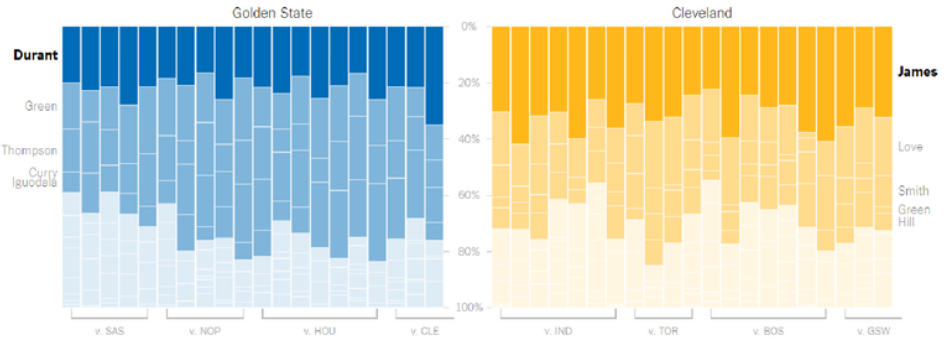

Figure 4.7 Players' Performance in Their Team

Over 40 years, there is actually no one who ever reach near the result that Lebron James scored. The figure below showed that the players who ever score the largest portion for their team's statistical performance throughout the playoffs. Kareem Abdul-Jabbar is the only player who consists more than 30 percent of his team's statistical performance.

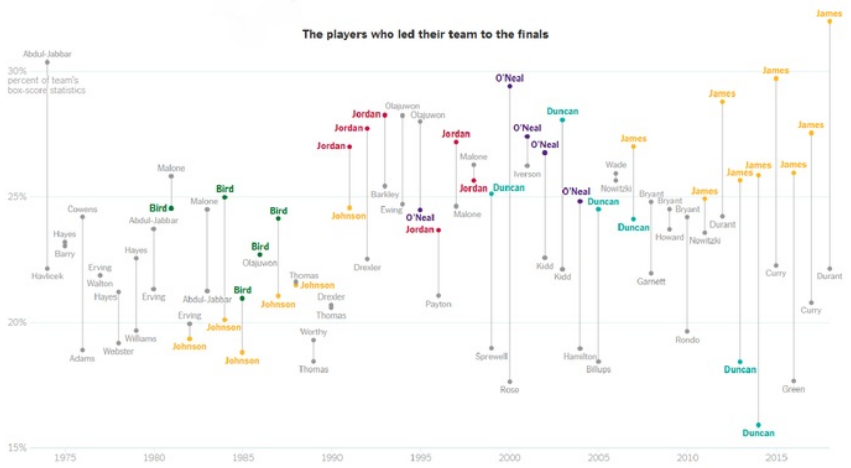

Figure 4.8 Players Outstanding Performance Percentage Who Led Their Team to Final

In 2016, James broke the record which led in all five categories for team's statistical data. He is the only man who ever claim the result like this. In the nine years he led the contribution portion with 6 times of four categories leading. In contrast, only Larry Bird and Tim Duncan led in four categories once.

2007 Cleveland		
PTS	501	28%
REB	161	19%
AST	159	45%
STL	34	24%
BLK	10	14%

2011 Miami		
PTS	497	26%
REB	176	20%
AST	123	36%
STL	35	23%
BLK	25	20%

2012 Miami		
PTS	697	31%
REB	224	24%
AST	129	32%
STL	43	26%
BLK	16	14%

2013 Miami		
PTS	596	27%
REB	193	22%
AST	152	32%
STL	41	21%
BLK	18	14%

2014 Miami		
PTS	548	28%
REB	141	21%
AST	95	26%
STL	36	24%
BLK	11	14%

2015 Cleveland		
PTS	601	30%
REB	226	24%
AST	169	47%
STL	33	26%
BLK	21	17%

2016 Cleveland		
PTS	552	25%
REB	200	22%
AST	160	36%
STL	49	30%
BLK	27	30%

2017 Cleveland		
PTS	591	28%
REB	164	22%
AST	141	36%
STL	35	26%
BLK	23	27%

2018 Cleveland		
PTS	725	34%
REB	193	22%
AST	190	47%
STL	30	24%
BLK	22	24%

Figure 4.9 Lebron James Performances in the Final

In conclusion, James is the significant player who can led his team not only to join the final and also the chance to get the champion.

4-3 Prediction of NBA Season 2018-19

NBA season of 2018-19 is still ongoing, however it only left 109 matches to finish the season game. Thus, the prediction had made to estimate which team will joined the Playoff in this season.

Team(East)	CurrentWin	UpdatedWin
Milwaukee Bucks*	**56**	**62**
Toronto Raptors*	**52**	**59**
Philadelphia 76ers*	**47**	**54**
Indiana Pacers*	**45**	**49**
Boston Celtics*	**44**	**49**
Brooklyn Nets*	**38**	**41**
Detroit Pistons*	**37**	**41**
Orlando Magic	37	38
Miami Heat*	**36**	**40**
Charlotte Hornets	35	39
Washington Wizards	31	34
Atlanta Hawks	27	29
Chicago Bulls	21	24
Cleveland Cavaliers	19	20
New York Knicks	14	17
**Team Joined Playoff*		

Team(West)	CurrentWin	UpdatedWin
Golden State Warriors*	**51**	**59**
Denver Nuggets*	**50**	**52**
Portland Trail Blazers*	**47**	**53**
Houston Rockets*	**47**	**49**
Utah Jazz*	**45**	**49**
Los Angeles Clippers*	**45**	**47**
Oklahoma City Thunder*	**44**	**48**
San Antonio Spurs*	**43**	**46**
Sacramento Kings	37	40
Minnesota Timberwolves	33	35
Los Angeles Lakers	33	37
New Orleans Pelicans	31	34
Memphis Grizzlies	30	32
Dallas Mavericks	29	32
Phoenix Suns	17	19
**Team Joined Playoff*		

Figure 4.10 Prediction of Playoff Qualification in Season 2018-19

In conclusion, this was only the bonus prediction for the future result of the Playoff qualification. The unfinished games had predicted and add to current game record, thus the top 8 of each conference had selected which the potential to join the Playoff of this season.

CHAPTER 5: CONCLUSION

From this analysis, there was clearly some issues occur in the current NBA system, especially in the Playoff sessions. Due to the stronger teams around the western conference, the parity problem occurred in the current NBA game, which make the top 16 teams in Playoff were not actually the strongest teams in the current season.

To begin the analysis for the result, the data was obtained from Basketball-Reference which is the database consists of all the statistical results in major basketball game. There are 24692 players' statistics from the collections with 51 attributes for each player. Selection had been made to pick the individual to the specific team to train. Different year's statistics of the individual were also taken for the modelling.

The reason of this analysis was using Naive Bayes for the supervised machine learning was because it does not need a lot of data to perform well. It needn't bother with a ton of information to perform well. It needs enough information to comprehend the probabilistic relationship of each feature in detachment with the result variable. Given that interactions between features were overlooked in the model, the instances of this connection are not required and so it need only less data compare to other algorithms, such as logistic regression. Furthermore, the overfit issue will not occur in this model even with only smaller sample size of data. After all, the accuracy of the model was quite impressive too which had 70.8% with only 2460 training data.

In conclusion, there was slightly different visually of the result with the actual one, elimination based and the round-robin rule. The comparison result clearly showed that, the western conference's team are generally stronger than the East which proven by the round-robin rule base and there were 5 teams of West and 3 teams from East which reached the top 8 placed. Thus, the analysis had concluded that, the parity problem did occur in NBA. To reduce the effect of the parity problem in NBA, this analysis was aim for raising the awareness of all the party and taking action to solve this phenomena in NBA which create the champion friendly environment for all the potential teams.

CHAPTER 6: RECOMMENDATION

Since the natural of the Naïve Bayes, each of the variables took as same weight during the analysis, so assume that the players' performance are independent depends on their position, with 5 of the starter, 5 from main bench team and other 3 as the bench's substitute, and each players contribute identical impact towards the result. Under this assumption, the model processed to get the result.

Due to the issue occur in the model, presented a general translation from Naïve Bayes into weighted model counting on conjunctive normal form (CNF). In particular, the methodology calls for encoding the probabilistic model, ordinarily a Bayesian system, as a propositional learning base in CNF with weights related to each model as indicated by the system parameters. Given CNF, registering the likelihood of some proof turns into a matter of summing the weights of all CNF models consistent with the proof. Various minor departure from this methodology have showed up in the writing as of late, that shift crosswise over three symmetrical measurements. The main measurement concerns the particular encoding used to change over a Bayesian system into a CNF. The second measurements identifies with whether weighted model checking is performed utilizing a hunt calculation on the CNF, or by ordering the CNF into a structure that renders weighted model counting a polytime task in the span of the assembled structure. The third measurement manages the local structures which are caught in the CNF encoding. (Chavira and Darwiche, 2008)

Thus, by applying the technique into the future model, it might slightly help the model to increase its accuracy, and also make the model seems more logical towards the actual result.

References

Thistimedoneliao

ORIGINALITY REPORT

12%	9%	4%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Tunku Abdul Rahman Student Paper	2%
2	Submitted to University of Newcastle upon Tyne Student Paper	1%
3	Submitted to University of Hong Kong Student Paper	1%
4	www.cs.waikato.ac.nz Internet Source	1%
5	www.psgminer.com Internet Source	1%
6	Submitted to Carnegie Mellon University Student Paper	1%
7	Submitted to Varsity College Student Paper	1%
8	Submitted to Palm Beach Currumbin State High School Student Paper	1%

9	en.wikipedia.org Internet Source	<1%
10	Submitted to American University of Beirut Student Paper	<1%
11	Submitted to Bridgepoint Education Student Paper	<1%
12	Submitted to University College London Student Paper	<1%
13	Submitted to Queen Mary and Westfield College Student Paper	<1%
14	Bert Fraussen, Timothy Graham, Darren R. Halpin. "Assessing the prominence of interest groups in parliament: a supervised machine learning approach", The Journal of Legislative Studies, 2018 Publication	<1%
15	Submitted to Universiti Teknologi MARA Student Paper	<1%
16	Mark Brown, Joel Sokol. "An Improved LRMC Method for NCAA Basketball Prediction", Journal of Quantitative Analysis in Sports, 2010 Publication	<1%
17	Submitted to University of Florida	

	Student Paper	<1%
18	Submitted to University of Surrey Student Paper	<1%
19	www.ijcst.com Internet Source	<1%
20	users.skynet.be Internet Source	<1%
21	link.springer.com Internet Source	<1%
22	eprints.utar.edu.my Internet Source	<1%
23	Qiubing Ren, Mingchao Li, Shuai Han. "Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives", Big Earth Data, 2019 Publication	<1%
24	www.citeulike.org Internet Source	<1%
25	digilib.its.ac.id Internet Source	<1%
26	rubasket.com Internet Source	<1%

27	go.mgoetze.net Internet Source	<1%
28	John A. David, R. Drew Pasteur, M. Saif Ahmad, Michael C. Janning. "NFL Prediction using Committees of Artificial Neural Networks", Journal of Quantitative Analysis in Sports, 2011 Publication	<1%
29	www.goldenstateofmind.com Internet Source	<1%
30	www.eecs.qmul.ac.uk Internet Source	<1%
31	mro.massey.ac.nz Internet Source	<1%
32	www.coolfactsforkids.com Internet Source	<1%
33	dynlab.mpe.nus.edu.sg Internet Source	<1%
34	www.assignmentmakers.com Internet Source	<1%
35	Submitted to Associatie K.U.Leuven Student Paper	<1%
36	www.budinst.gov.kh Internet Source	<1%

37	psasir.upm.edu.my Internet Source	<1%

Exclude quotes	On	Exclude matches	< 8 words
Exclude bibliography	On		