Finding the Cause of the Unexpected Slump of Kante and Chelsea FC Based on Regression Analysis

Abstract

N'golo Kante, Chelsea's most valuable midfielder in the 2018/2019 season, led the team to a Europa League victory. However, unexpectedly, in the 2019/2020 season, Chelsea's winning rate dropped when Kante played. This phenomenon lasted for the whole season. The goal of the research is to find the hidden causes of this unexpected phenomenon. To find the answer, statistics related to N'golo Kante and Chelsea FC were collected from various sources and analyzed using interdisciplinary methods. Automated web crawling based on Python Selenium Webdriver is used for collecting data, and machine-learning based on linear and polynomial regression models is mainly used for analysis. The analysis indicates that Kante's changed role and teammates, injuries, and poor relationship with his manager harmed both his performance and Chelsea FC in the 2019/2020 season. To overcome their slump and keep a positive relationship with their star player, Chelsea FC needs to manage Kante in a special way so that he can function smoothly and perform to his true ability.

Bugil Academy
Wonho Lim
wonholim02@gmail.com

I. Introduction

In the 2018-2019 season, when the Chelsea Football Club of England Premier League won UEFA Europa League, N'golo Kante was one of the key players of Chelsea FC. Kante was chosen as a member of Team of The Year by FIFA in both the 2018 and 2019 season. He showed an impressive performance, and many fans and experts praised Kante. However, in the 2019-2020 season, he and his team experienced a terrible unexpected slump. Starting from 4-0 defeat at Old Trafford, both Kante and Chelsea FC experienced consecutive failures and hardships. Kante had been playing for Chelsea FC since 2016, but that situation had never happened, and that lasted for a long time. For both Chelsea FC and Kante, the first half of the season was disastrous. However, no research mainly discussed the unexpected situation. This research aims to find hidden reasons why this unexpected problem has occurred, based on a complex analysis of various statistics and information.

In "A study on KBO league foreign pitchers' re-sign possibilities using decision tree analysis,"[1] the researcher uses Decision Tree and Random Forest, which are machine learning models, to do research about which record increases the contact extension probability of foreigner pitchers of KBO. The researcher collected various pitcher records to make a model so that the random forest can have decision trees that have a small error range. The result shows that ERA is the most significant factor that affects the contract extension.

---

[1] Gyuin Hwang (2018). "A study on KBO league foreign pitchers' re-sign possibilities using decision tree analysis." Master Thesis, Korea Cyber University graduate school of interdisciplinary information studies.

"DATA ANALYTICS IN SPORTS: IMPROVING THE ACCURACY OF NFL DRAFT SELECTION USING SUPERVISED LEARNING"[2] uses various Supervised Learning Classification methods and algorithms to make more accurate NFL draft result prediction. The researcher trained Naive Bayes, Multiplayer Perception, Logistic Regression, RBF Network with the draft result and the post year record. Then, the researcher checked the accuracy and predicted the outcome. In conclusion, the researcher found that machine learning can be used to predict the probability of the picked player's success in the NFL.

II. Method

In this research, various statistics and informational data from various reliable internet websites and articles are collected for the analysis. For the analysis of a broad range of factors, data are collected from various websites. Websites like the Official England Premier League website[3][4][5], Transfermarkt[6][7][8], and other reliable sources and websites are mainly used.

Two methods are mainly used for collecting data. The first method is using "automated web crawling" by using Jupyter Notebook based on Python 3.0. Selenium Webdriver, which is generally used for browser controlling, is used as a base of web

---

[2] Gary McKenzie (2015). DATA ANALYTICS IN SPORTS: IMPROVING THE ACCURACY OF NFL DRAFT SELECTION USING SUPERVISED LEARNING. In Partial Fulfillment of the Requirements for the Degree Master of Science, University of Missouri-Columbia.

[3] "Chelsea FC Statistics: Premier League." *FC Statistics | Premier League*, www.premierleague.com/clubs/4/Chelsea/stats

[4] "N'Golo Kanté Statistics: Premier League." *Statistics | Premier League*, www.premierleague.com/players/13492/N%27Golo-Kant%C3%A9/stats

[5] *Premier League Table, Form Guide & Season Archives*, www.premierleague.com/tables

[6] *N'Golo Kanté - Player Profile 19/20.* www.transfermarkt.com/ngolo-kante/profil/spieler/225083

[7] "Frank Lampard - Player Profile." *Transfermarkt*, www.transfermarkt.com/frank-lampard/profil/spieler/3163

[8] "Maurizio Sarri - Manager Profile." *Transfermarkt*, www.transfermarkt.com/maurizio-sarri/profil/trainer/10073

crawling. Chrome was mainly used for the main web browser. This automated method is very convenient for dealing with a significant amount of data more efficiently as it only requires coding and time. However, because it is not efficient for collecting a small amount of data as it requires coding, I used another method. The second method is collecting data by bare hand, which process is very simple but slow. This method was efficient for collecting a small amount of data or literal data. For example, changed teammates are collected by bare hands.

2018-2019 season statistics and 2019-2020 statistics of Chelsea FC, Kante, Kante's teammates, Frank Lampard(Manager), Maurizio Sarri(former Manager), are mainly collected. Other useful information from different seasons and leagues are broadly collected individually.

Statistic data from the 2018-2019 and 2019-2020 season of EPL, Chelsea FC, and N'golo Kante is used for mathematical and statistical analysis. Various aspects of data, including difference, similarity, change, trend, and specialty, are analyzed in different ways for the analysis's accuracy.

Machine learning is used for the analysis based on a linear regression model and polynomial regression model for analysis and prediction. For examination of the reliability of the model, a coefficient of determination is used. Linear regression model and Polynomial regression model were generated using both Python 3.0 and Microsoft Excel, which are generally used for machine learning and data analysis. The "Forecast" function, which predicts the expected result using the existing data and analysis model, is used for further analysis and prediction, especially in Excel. Moreover, non-statistical evidence and

information that cannot be examined by computer are also analyzed. Finally, for more complex understanding and analysis, new indicators not usually provided are newly created in this research.
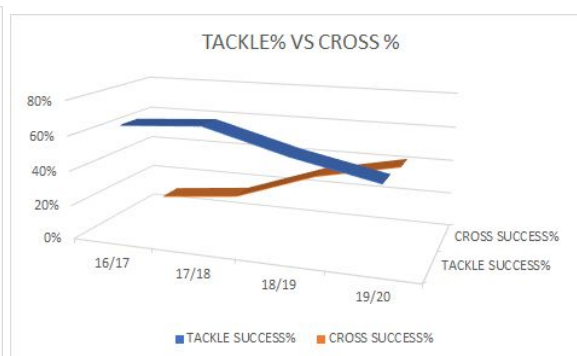
III.Result

Factor 1: Position & Playstyle

Inthe 2019-2020 season, Kante showed a much more aggressive playstyle than in the 2018-2019 season. The primary role of Kante had changed from defending and tackling to scoring more goals and making crosses. Considering the fact that Kante was known as one of the best defensive center midfielders in the world, his changed aggressive playstyle is a notable factor.[9] Several statistical evidence shows that Kante's playstyle has changed in many aspects, resulting in negative effects. In this section, for better comparison, the 2019-2020 season's records, except percentage and ratio type records, are multiplied by two because the 2018-2019 record is from 36 games and 2019-2020 records are from 18 games.
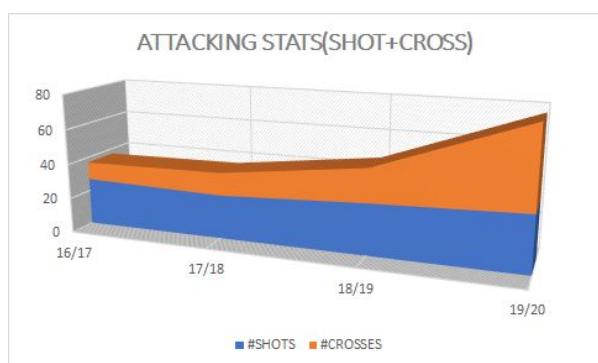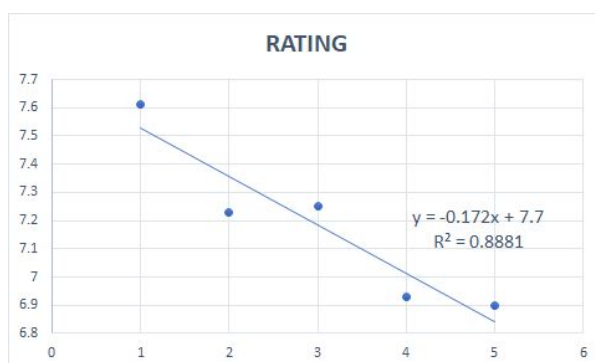

Attacking Stats(Shot+Cross) - (Picture 1)   Tackle% vs Cross % Change - (Picture 2)

---

[9] Matt Gunn. "N'Golo Kanté PoTY Analysis." *ESDF Analysis*, 18 May 2017, www.esdfanalysis.com/player-analysis/ngolo-kante-poty-analysis/

Recent 5 Year Rating - (Picture 3)

The Attacking Stats(Shot + Cross) graph shows that Kante's attacking record became more aggressive as time went on since he came to Chelsea FC. Rather than passing safely, he tried to shoot or cross the ball in the 2019-2020 season. Moreover, while the tackle success percentage has gradually decreased by about 20 percent, the cross success percentage has gradually increased by about 20 %. The point is, Kante's rating has declined rapidly as he became more and more aggressive. The linear regression model shows that Kante's rating has decreased overall for the last five years with the slope of -0.172, which is a significant number, while the attacking record increased simultaneously. The coefficient of determination is 0.88, showing that the model is reasonably reliable. These graphs prove that Kante became much more aggressive in the 2019-2020 season, and that lowered Kante's value.
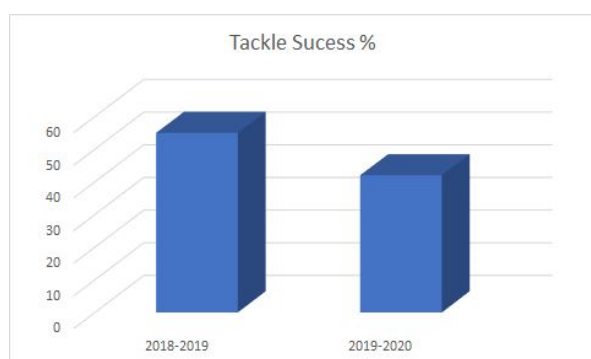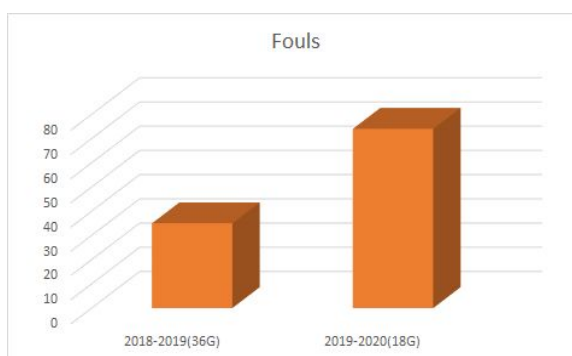

Tackle Success % - (Picture 4)
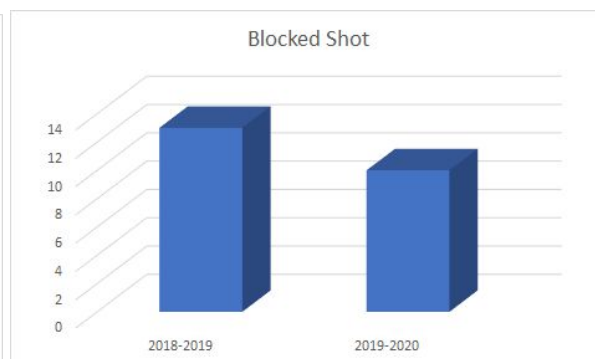
Blocked Shots - (Picture 5)
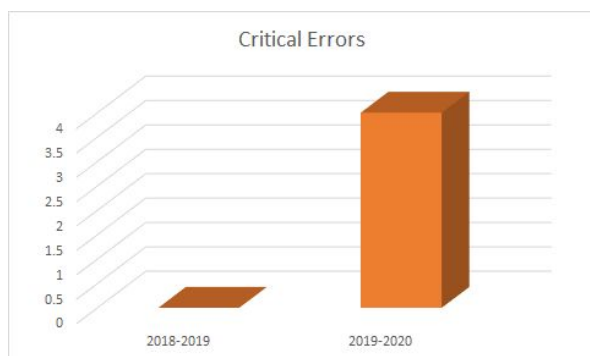
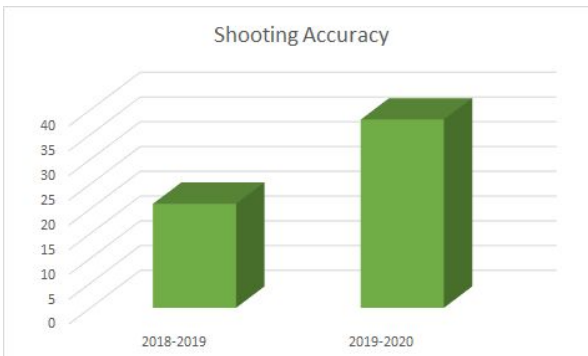# Critical Errors - (Picture 6)   # Fouls - (Picture 7)

In the 2018-2019 season, Kante contributed a lot to the team's defense. However, as the graph shows, tackle success percentage and the number of blocked shots decreased rapidly in the 2019-2020 season. Also, the number of critical errors leading to a goal, which he recorded zero in the 2018-2019 season, was increased by 4. The number of fouls, which can disturb the team's build-up process or give the opposition a free chance, also rapidly increased. Because other Chelsea FC's defenders also showed poor performance in 2019-2020, Kante's poor defense performance resulted in more significant problems. For instance, because Chelsea FC's main concern was set-piece defense in the 2019-2020 season, an increased number of fouls made the team suffer. This evidence proves that Kante showed low defensive ability in the 2019-2020 season, making it hard for the team to block the goal.


Goal/Match - (Picture 8)   Shooting Accuracy - (Picture 9)

While Kante's defensive ability rapidly decreased, he showed better performance as an attacker in several areas. Mainly, he showed much better performance for the shooting.

Goal/Match and shooting accuracy increased rapidly in the 2019-2020 season. However,

considering the fact that Chelsea FC's strategy changed from defensive Sarri Ball to

intensive attacking Lampard Ball, it is not special to see the increased Goal/Match rate.

Lampard required players to shoot more. As a result, Chelsea FC itself scored more goals this

season, but it lost more goals than that. In other words, it was not effective for Kante to shoot

more not only because his accuracy is still low compared to other attackers but also because

his team lost more points by putting Kante in front. Moreover, other statistics show that the

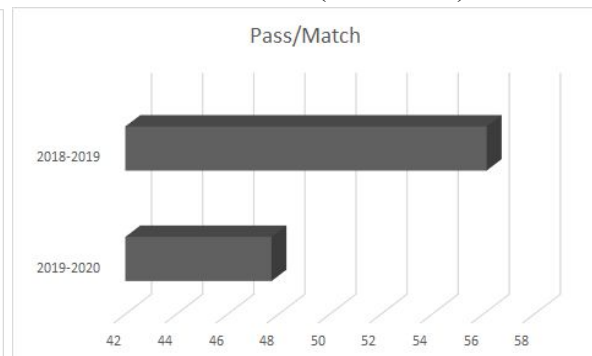increased number of attacks was not that meaningful.
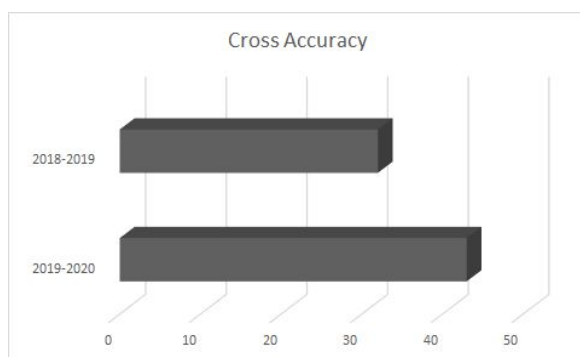
Cross Accuracy- (Picture 10)
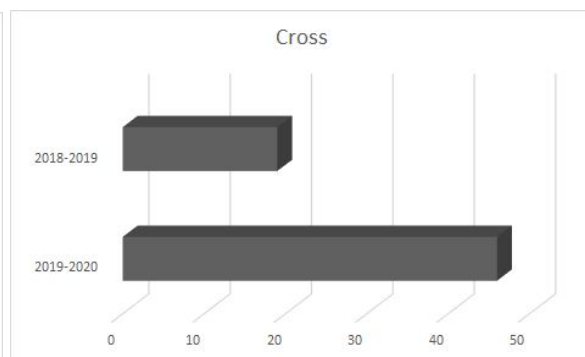
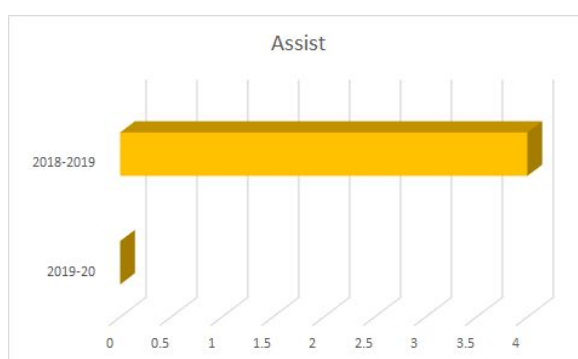# Cross - (Picture 11)
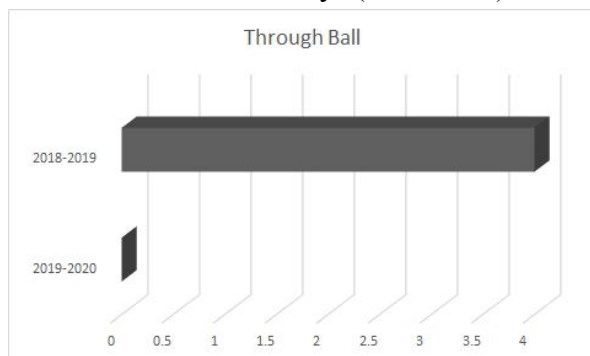
Through Ball - (Picture 12)

Pass/Match - (Picture 13)

The number of crosses and cross accuracy also increased, as the graph above shows.

But the point is that the number of keys through ball pass and pass/match rapidly decreased.

In other words, while Chelsea FC made their attacking process difficult, they also lowered the

possibility of making a great chance. Kante was the one who contributed a lot to the team's

build-up process in the back half of the field, but the number of passes per game notably

decreased while crosses increased. As a result, Chelsea FC had trouble playing the ball to the other side for half of the season. Also, Kante was a player who prefers low and short passes rather than long crosses, but the style was drastically changed, making Kante take time to be adopted.
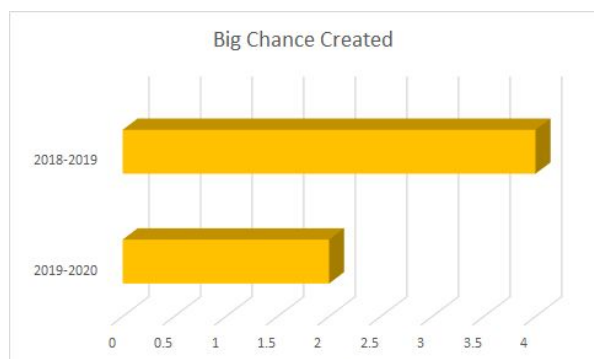
Big Chance Created (Picture 14)   # Assist - (Picture 15)

Moreover, Kante made many crosses than before but did not have any assistance in the first half of the season. Also, the number of big chances created decreased by half. This shows that Kante's aggressive playstyle was not that effective and meaningful for the attacking process. In other words, Kante's aggressive playstyle made it harder for Kante to contribute to the team's overall attacking process even though he gave up many defensive roles. Kante's changed playstyle of doing long balls and concentrating on shooting rather than blocking shots negatively affected Kante and the team.


2018-2019 Position - (Picture 16)   2019-2020 Position - (Picture 17)

Along with the change in his playstyle, his role in the team also changed as the team's manager changed in the 2019-2020 season. In the 2019-2020 season, Frank Lampard used an aggressive Gegenpressing tactic with the formation of 4-2-3-1 or 4-1-2-3. In Frank Lampard's team, Kante sometimes played in the center midfielder and sometimes in the center defensive midfielder. However, in the 2018-2019 season, Maurizio Sarri used a ball possessing strategy with the formation of 4-3-3. Especially, Sarri frequently put Kante only in the right-center box to box center midfielder in this season. As long as Kante played only as a CM for the whole season in 2018-2019, adapting to the new formation and the new position(CDM) would be hard in the 2019-2020 season.
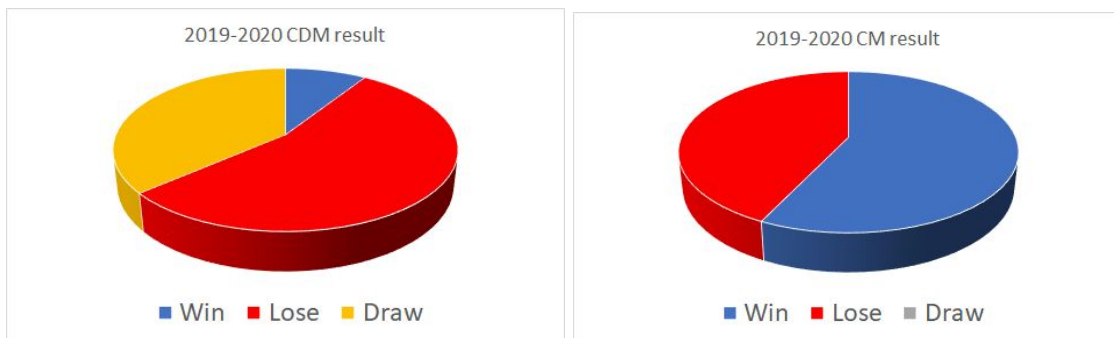
2019-2020 CDM Result - (Picture 18)   2019-2020 CM Result - (Picture 19)
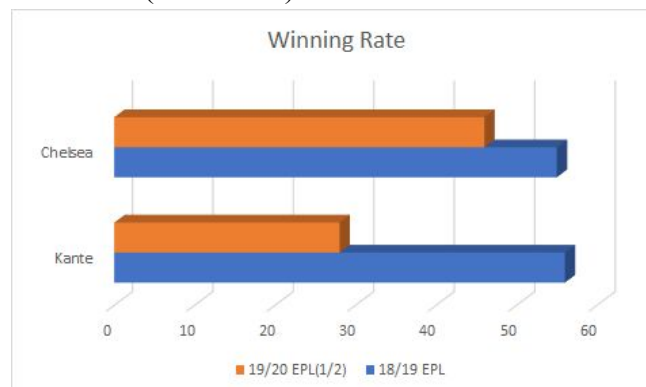
Winning Rate - (Picture 20)

The statistics prove that Kante had a hard time in the new position. In the 2019-2020 season, when he played as a center defensive midfielder, the winning rate was very low(1W 6L, 14%), compared to his winning percentage as a CM. As it is already mentioned, his playstyle became much more aggressive under Lampard's management. However, what

happened to him is that he had to play as a defensive midfielder while he was changing his playstyle more aggressively. It is evident that the "aggressive" style does not fit the "defensive" center midfielder position. This complicated situation made Kante struggle to adapt to his new position, develop his playstyle, and find his role in the team. Therefore, Kante's new CDM position in a new squad, new formation, and new strategy made Kante have a hard time playing as usual.

Factor 2: Teammates & Team as a whole

| IN +	OUT -
Christian Pulisic(Transfer, 64M€)	Eden Hazard(Transfer, 100M€)
Kurt Zouma(Loan Back)	David Luiz(Transfer, 8M£)
Fikayo Tomori(Loan Back)	Alvaro Morata (Loan+Transfer, 65M€)
Tammy Abraham(Loan Back)	Cesc Fabregas(Transfer, 10M€)
Mason Mount(Loan Back)	Gonzalo Higuain(Loan Back)
	Gary Cahill(Transfer, FA)
	Danny Drinkwater(Loan)

2019/2020 Transfer Result - (Picture 21)

At the beginning of the 2019-2020 season, many players who played a lot of time with Kante left Chelsea FC. Eden Hazard, David Luiz, Cesc Fabregas, Gonzalo Higuain, Alvaro Morata, Danny Drinkwater, and Gary Cahill left the first team. Eden Hazard, David Luiz, Cesc Fabregas, Alvaro Morata, and Gary Cahill were included in the list of Top 10 players who Kante played the most in his career. Teamwork is an essential part of soccer, as eleven players play together in the field. However, Kante eventually lost a significant number of teammates who he collaborated with for a long time. Because of this change, Kante had to adapt to a new style team with new members. Moreover, losing several valuable players
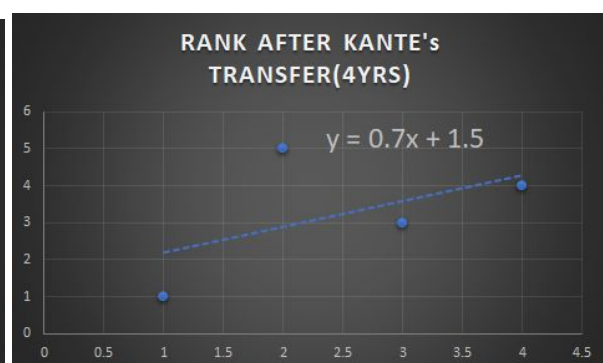
damaged Chelsea FC's overall team ability, making both Kante the whole team suffer. Especially Eden Hazard, who was the most valuable player in Chelsea FC in the 2018-2019 season, also left the team. Statistically, Chelsea FC recorded (21W/9D/8L/72Pts - 3rd place) in the 2018-2019 season while they recorded (20W/6D/12L/66Pts - 4th place) in the 2019-2020 season.

Moreover, in the 2019-2020 season, players who played a lot of time with Kante showed a much lesser appearance in the previous year. Frank Lampard, the new manager, did not prefer several players who Maurizio Sarri frequently employed in the 2018-2019 season. Olivier Giroud, Pedro Rodriguez, Ruben Loftus-Cheek, Marcos Alonso, and Emerson Palmieri showed a much lesser appearance than the 2018-2019 season. However, in the 2019-2020 season, Frank Lampard used many new players who are mostly young. Christian Pulisic, Mason Mount, Reece James, Fikayo Tomori, Kurt Zouma, Tammy Abraham, and Michy Batshuayi, who Kante never played together last season, joined the squad in the 2019-2020 season. Especially Tammy Abraham, Kurt Zouma, Fikayo Tomori, Mason Mount, and Christian Pulisic are frequently included in the starting lineup of Chelsea F.C during the first half of the season. These new teammates made it hard for Kante to keep his original role and maintain his ability.
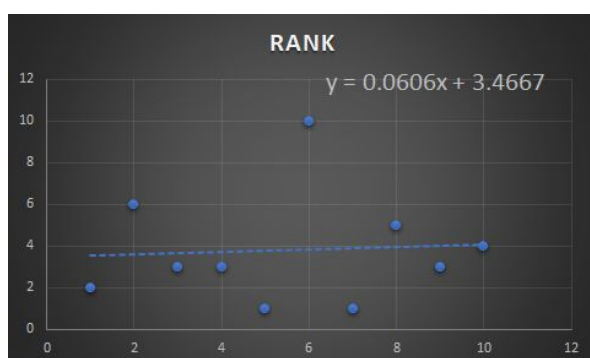
Rank Change - (Picture 22)   Rank Change After Kante- (Picture 23)

Finally, as a team, Chelsea FC has been weakened gradually for several years regardless of Kante's ability, making it harder for Kante to show better work. The linear regression model for the last ten years rank of Chelsea FC shows that the numerical rank of Chelsea FC increased for the previous ten years overall, which means lower rank. Moreover, a linear regression model for rank after Kante's transfer shows that the team's rank fell even rapidly after the transfer, with a linear slope of 0.7, which is a notable change rate. Additionally, the linear regression model predicted that next year's Chelsea FC's rank would be about 4.13 based on a 10-year rank trend. Therefore, Kante and the team's slump has lasted for recent years, making Kante have a hard time fully showing his ability.

Factor 3: Injuries and Physical Problems.

Kante has a height of 168 cm and a weight of 70 Kg. The average height of England Premier League midfielders in 2019 was approximately 181cm. Considering that fact, Kante is regarded as a relatively short player. Most successful short midfielders, such as Alan Ball Jr and Billy Bremner, had excellent dribbling and passing skills rather than physical ability. Speed, pass, and dribbling skills are important for most short midfielders because they cannot win the ball through physical competition against tall and heavy players. However, Kante faced consecutive injuries related to his legs in the 2019-2020 season. Leg injuries can lead to critical problems for any soccer player. Especially, leg problems frequently cause problems in running and dribbling. Kante would also have suffered from the same pain because he has to move a lot, as usual, to contribute to the team, but he cannot because of his consecutive injury. Kante's overall ability and contribution to the team might have decreased rapidly in the 2019-2020 season because of consecutive leg related injuries.

Injury Information

-----------------------------Before 2019-2020 Season-----------------------------

May 6, 2019 Hamstring 22 Injury days.

Jun 3, 2019 Knee Injury 9 days.

-----------------------------During 2019-2020 Season-----------------------------
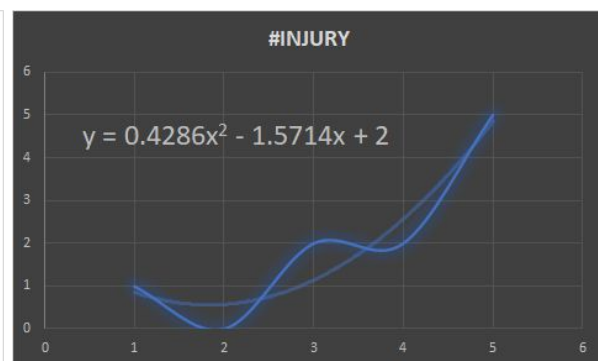
Aug 19, 2019 Ankle Injury 33 days. - Period 1

Oct 15, 2019 Unknown Injury 20 days. - Period 2
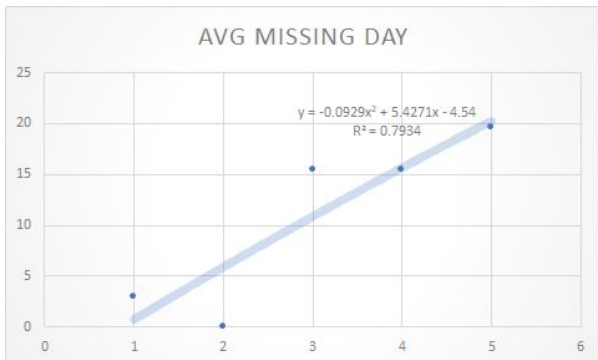
Feb 18, 2019 Muscle Injury 21 days. - Period 3

Injury 1 - (Picture 24)

$y = 0.4286x^2 - 1.5714x + 2$

Injury 2 - (Picture 25)

$y = 7.7857x^2 - 24.614x + 20.8$
$R^2 = 0.9215$

Missing Day - (Picture 26)

$y = -0.0929x^2 + 5.4271x - 4.54$
$R^2 = 0.7934$
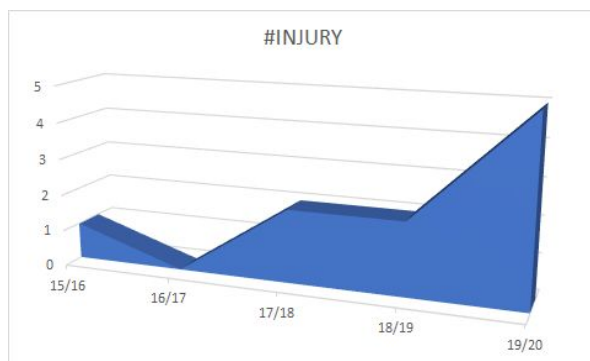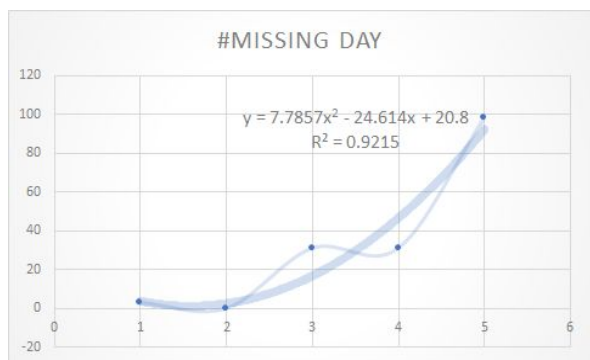
AVG Missing Day- (Picture 27)

The two #INJURY graph shows that Kante got more and more injuries as the year

went by. According to the polynomial regression model, the number of missing days due to

injury since Kante moved to EPL has increased almost exponentially. The polynomial

regression model predicted that Kante's average missing day per injury in the sixth

season(2020-2021) to be 25.33 days. Moreover, each season's average missing day per injury

has increased almost linearly with the slope roundabout 5 in the same period. The coefficient

of determination was 0.9215 and 0.7934 for each graph, making it evident that the

seriousness of injury got heavier. This polynomial regression analysis proves that Kante's

severity of injury had increased for the last five years. Because his primary position is Box to

Box midfielder, which means he moved a lot every game, his leg got a lot of stress for five

years. In conclusion, Kante's physical condition got worse gradually and struck Kante

critically in the 2019-2020 season, and an even worse injury is expected if the situation does

not change in the next season.
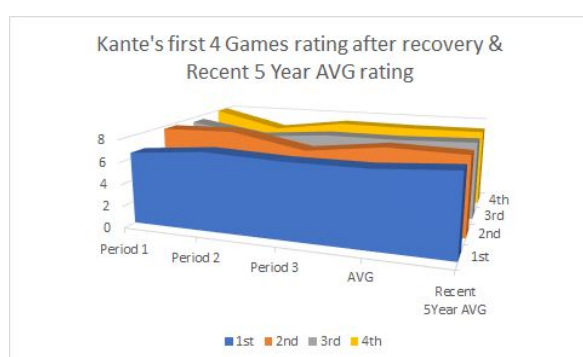
First 4 Games Rating 1 - (Picture 28)   First 4 Games Rating 2 - (Picture 29)
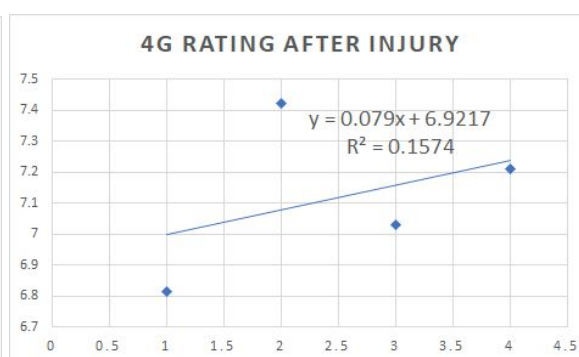
These visual graphs show that Kante was gradually adapted to the team after each

injury. The first game rating of all three after-injury periods was always lower than the recent

five-year average rating. Moreover, in periods 3 and 4, the ratings of both the 3rd and 4th

games were also lower than the recent 5-year average rating. The linear regression model for

the first four-game average rating after injury again proves that Kante showed better

performance as the number of games he played increased. Even though the coefficient of

determination is 0.1574, which is not highly accurate, considering that the size of data was

not big, the linear regression model is considered evidence as it could be used to check

general trends. The linear regression equation of rating shows the positive slope of 0.079,

proving that Kante recovered gradually. Kante's frequent injury negatively affected Kante's

primary ability while making him suffer to adapt to the team again after injury. Therefore, injury can be considered one of the significant factors that affected Kante's ability in the 2019-2020 season.

IV.Conclusion

Complex analysis of Kante and Chelsea FC proved that a number of significant factors affected Kante and the team's 2019-2020 season performance negatively. The notable factors were changed position and role, unsuitable play style under a different manager, changed teammates, and consecutive injuries. Moreover, the team's seasonal slump made Kante harder to play as usual. In conclusion, in the 2019-2020 season, various combinational factors created the slump of both N'golo Kante and Chelsea FC.

Every player experiences slump at least once in their life, and N'golo Kante is not the exception. This research shows that Kante has significant problems that need some adjustment. In other words, waiting until the situation is better will not help both the team and the player. It is necessary for Chelsea FC and Kante to make a change to overcome this situation. The coaches and Kante should discuss and find the right position and play style for Kante for his performance. Moreover, Kante needs proper care as he suffered from consecutive injuries. As the machine learning prediction based on the linear regression model and polynomial regression model shows, the more consecutive injury might affect him even more. Through long-time management, Kante and Chelsea FC could change the situation.

Later, several additional analyses can be done to develop research in various ways. Even though the machine learning model using linear regression and polynomial regression

analyzed the statistics' general trends, more complex models can be used for more accurate

analysis. Moreover, because only a recent five-year record of N'golo Kante, which is

relatively small, is used for research, by using extra data from another year or other players, a

more complex understanding could be made. Analyzing another factor like weather and the

number of spectators can also improve the research.

V.References

1. Gyuin Hwang (2018). "A study on KBO league foreign pitchers' re-sign possibilities using decision tree analysis." Master Thesis, Korea Cyber University graduate school of interdisciplinary information studies.
2. Gary McKenzie (2015). DATA ANALYTICS IN SPORTS: IMPROVING THE ACCURACY OF NFL DRAFT SELECTION USING SUPERVISED LEARNING. In Partial Fulfillment of the Requirements for the Degree Master of Science, University of Missouri-Columbia.
3. "Chelsea FC Statistics: Premier League." *FC Statistics | Premier League*, www.premierleague.com/clubs/4/Chelsea/stats
4. "N'Golo Kanté Statistics: Premier League." *Statistics | Premier League*, www.premierleague.com/players/13492/N%27Golo-Kant%C3%A9/stats
5. *Premier League Table, Form Guide & Season Archives*, www.premierleague.com/tables
6. *N'Golo Kanté - Player Profile 19/20.* www.transfermarkt.com/ngolo-kante/profil/spieler/225083
7. "Frank Lampard - Player Profile." *Transfermarkt*, www.transfermarkt.com/frank-lampard/profil/spieler/3163
8. "Maurizio Sarri - Manager Profile." *Transfermarkt*, www.transfermarkt.com/maurizio-sarri/profil/trainer/10073
9. Matt Gunn. "N'Golo Kanté PoTY Analysis." *ESDF Analysis*, 18 May 2017, www.esdfanalysis.com/player-analysis/ngolo-kante-poty-analysis/
10. Scott, Lee. "Frank Lampard at Chelsea 2019/20 - Tactical Analysis." *Total Football Analysis Magazine*, 4 July 2019, totalfootballanalysis.com/head-coach-analysis/frank-lampard-chelsea-2019-20-tactical-analysis
11. Gerry Crisandy"Frank Lampard Speaks about N'Golo Kante's Position Change and Midfield Competition." *The Chelsea Chronicle*, 30 June 2020, www.thechelseachronicle.com/club-news/frank-lampard-speaks-about-ngolo-kantes-position-change-and-midfield-competition/
12. Kostanjšak, Domagoj. "Maurizio Sarri Tactical Analysis: Sarriball at Chelsea vs Sarriball at Napoli." *Total Football Analysis Magazine*, 31 Oct. 2019,

totalfootballanalysis.com/head-coach-analysis/maurizio-sarri-chelsea-napoli-head-coach-tactical-analysis-statistics

13. White, Mark. "Why N'Golo Kante Is No Longer Chelsea's Most Important Midfielder." *Fourfourtwo.com*, FourFourTwo, 26 June 2020, www.fourfourtwo.com/features/why-ngolo-kante-is-no-longer-chelseas-most-important-midfielder