Report

Kaiyu Yan

December 4, 2018

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Abstract

Introduction

Background

Trying to determine who was the best quarterback or running back through out the year is usually an interesting topic for media and college football fans. Also, to win a game, coach always need to put his best players on the field. Therefore, identifying and projecting the performance for players is particularly important for a team. People like to take one of the most basic statistic to measure performance. For example, yards per attempt(YPA), we like to use this measurement to determine the performance of a quarterback or running back's performance. However, for most time, when we look at the leaders in yards per attempt, we will notice that the statistical data is not useful. Because the highest yards per attempt always dependent on the lowest number of attempts as shown in table 1.1. Therefore, for this project, we are trying to build a model that will generalize the most unbiased information to help us to determine the best performed player based on certain measurement.

```
img1_path <- "picture1.PNG"
img1 <- readPNG(img1_path, native = TRUE, info = TRUE)
# Small fig.width
include_graphics(img1_path)
```

Data and Method

Data Source

Model Used

EDA and Result

EDA

Model Choice

Interpretation

Model Checking

Discussion

Limitation

Future Direction

```
library(tidyverse)
Data_2012 <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/team-game-statistics.csv")

Team <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/team.csv")
team_defense <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/team-defense.csv")
team_efficient <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/team-efficient.csv")
power_index <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/power-index.csv")
schedule <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/game.csv")
Player_rush <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/rush.csv")
player <- read.csv("C:/Users/Mark/Desktop/cfbstats-com-2012-1-5-4/player.csv")

rushing_stats <- aggregate(Data_2012[,3:4],list(Data_2012$Team.Code),sum)
colnames(rushing_stats)[1] <- "Team.Code"
rushing_stats <- inner_join(rushing_stats,Team,by = "Team.Code")
rushing_stats <- mutate(rushing_stats,YPC = Rush.Yard/Rush.Att)
colnames(Team)[2] <- "TEAM"
colnames(rushing_stats)[4] <- "TEAM"
A <- inner_join(power_index,Team,by = "TEAM")
A <- inner_join(A,team_efficient,by = "TEAM")
A <- inner_join(A,team_defense,by="TEAM")
A <- inner_join(A,rushing_stats,by="TEAM")
colnames(schedule)[3] <- "Team.Code"
schedule <- inner_join(schedule,Team,by = "Team.Code")
colnames(schedule)[4] <- "Team.Code1"
colnames(Team)[1] <- "Team.Code1"
schedule <- inner_join(schedule,Team,by = "Team.Code1")
schedule <- select(schedule,1,7,9)
colnames(schedule)[2] <- "Home Team"
colnames(schedule)[3] <- "Away Team"
```

```
dat <- as.data.frame(NULL)
for (i in 1:93){
  A1 <- filter(schedule,schedule$`Home Team` == A[i,2])
  A2 <- filter(schedule,schedule$`Away Team` == A[i,2])
  A2 <- A2[c(1,3,2)]
  colnames(A2)[2] <- "Home Team"
colnames(A2)[3] <- "Away Team"
  B <- rbind(A1,A2)
  dat <- rbind(dat,B)
}
colnames(dat)[3] <- "TEAM"
dat <- inner_join(dat,team_defense,by="TEAM")
dat <- inner_join(dat,power_index,by="TEAM")
dat <- inner_join(dat,team_efficient,by="TEAM")
dat <- select(dat,1:8,10,13,15)
Oppo_defense <- aggregate(dat[,5:11],list(dat$`Home Team`),mean)
colnames(Oppo_defense)[2] <- "Opp.Att.Allowed"
colnames(Oppo_defense)[3] <- "Opp.Yds.Allowed"
colnames(Oppo_defense)[4] <- "Opp.Ypc.Allowed"
colnames(Oppo_defense)[5] <- "Opp.Ypg.Allowed"
colnames(Oppo_defense)[1] <- "TEAM"
colnames(Oppo_defense)[6] <- "Opp.FPI"
colnames(Oppo_defense)[7] <- "Opp.Def.Eff"
colnames(Oppo_defense)[8] <- "Opp.Overall.Eff"


Team_stats <- select(A,2,3,7,8,10,17,18,20)
Team_stats <- inner_join(Team_stats,Oppo_defense,by = "TEAM")

Player_rush <- aggregate(Player_rush[,5:6],list(Player_rush$Player.Code),sum)
Player_rush <- mutate(Player_rush,YPA = Yards/Attempt)
player$Fullname <- as.character(paste(player$First.Name,player$Last.Name, sep = " "))
player <- select(player,1,2,7,14)
colnames(Player_rush)[1] <- "Player.Code"
Player_rush <- inner_join(Player_rush,player,by="Player.Code")
# Player_rush <- filter(Player_rush,Position == "RB")
Player_rush$Position <- as.character(Player_rush$Position)
 Player_rush$Position[Player_rush$Position == ""] <- NA
 Player_rush$Attempt[Player_rush$Attempt<1]  <- NA
 Player_rush <- na.omit(Player_rush)
Player_rush <- select(Player_rush,-1)
colnames(Player_rush)[4] <- "Team.Code1"
Player_rush <- inner_join(Player_rush,Team,by ="Team.Code1" )
Player_rush <- select(Player_rush,-4,-8)
Player_rush <- inner_join(Player_rush,Team_stats,by = "TEAM")

DT::datatable(Player_rush,
          filter = 'top', options = list(
            pageLength = 12, autoWidth = TRUE
          ))
ggplot(data = Player_rush)+
  geom_point(mapping = aes(x=Attempt,y = YPA))
```

```
ggplot(data = Player_rush) +
  aes(x = YPA) +
  geom_histogram(bins = 50, fill = "#0c4c8a") +
  theme_classic()

ggplot(data = Player_rush) +
  aes(x = FPI, y = YPA) +
  geom_point(color = "#0c4c8a") +
  geom_smooth(se = F) +
  theme_classic()

ggplot(data = Player_rush) +
  aes(x = Opp.FPI, y = YPA) +
  geom_point(color = "#0c4c8a") +
  geom_smooth(se = F) +
  theme_classic()

ggplot(data = Player_rush) +
  aes(x = Opp.FPI, y = Opp.Yds.Allowed) +
  geom_point(color = "#0c4c8a") +
  geom_smooth(se = F) +
  theme_classic()



library(nlme)
m1 <- lme(fixed = YPA~OFFENSE+Rush.Att+YPC+Opp.Ypc.Allowed+Opp.FPI,random = ~1|TEAM,data=Player_rush)
summary(m1)
library(lme4)
m2 <- lmer(YPA~ FPI+ (1|TEAM),Player_rush)
summary(m2)
head(fixef(m2))
head(ranef(m2)$TEAM)
head(coef(m2)$TEAM)
```