```
---
title: "Classification"
output:
  pdf_document: default
  html_document: default
date: "2023-02-18"
name: Imad Siddiqui
---
```

```{r}
set.seed(3)
i <- sample(1:nrow(male_teams), 0.8*nrow(male_teams), replace=FALSE)
train <- male_teams[i,]
test <- male_teams[-i,]

```

```{r}
mean(train$overall)
range(train$overall)
mean(train$midfield)
mean(train$attack)
mean(train$defence)

```

```{r}
plot(train$overall~train$attack, xlab="attack", ylab="overall")
abline(lm(train$overall~train$attack), col="red")

plot(train$overall~train$defence, xlab="defence", ylab="overall")
abline(lm(train$overall~train$defence), col="blue")
```

In this cell, I built a logistic regression model using attack, defense, and midfield as predictors for overall.

```{r}
glm1 <- glm(train$overall~train$attack+train$defence+train$midfield)
summary(glm1)

```

To summarize, since the residual deviance is much lower than the null, the predictors definitely played a big role in accurately predicting the team's overall performance.

In this cell, I tried installing the library required for naive bayes, but I could not for some reason. For that reason, I could not finish this assignment.

```{r}
library(e1071)
nb1 <- naiveBayes(train$overall~train$attack+train$defence+train$midfield)
nb1

```

To summarize, logistic regression is a good way to find out how well you are able to predict qualitative data. It's easy to train data using this algorithm. The biggest disadvantage is that it assumes linearity just like with linear regression. This will lead to high bias. Naive bayes will solve this problem but is harder to work with.