2011-2012 English Premier League Goal Analysis

# Introduction and Data Description

The following analysis was performed on data provided by Opta on the 2011 - 2012 English Premier League (EPL) season. This link will take you to the data: <https://datahub.io/dataset/uk-premier-league-match-by-match-2011-2012>.

Each row of data represents statistics on a single player from one game of the 2011 - 2012 EPL season. 201 different statistics were collected for each player for every game they played in during the season. The other 9 fields were data about the player, the team they played on, the team they were playing against, and the date of the game.

# Exploratory Analyis

I started by importing the data.

epl <- read.csv("epl.11-12.csv", na.strings = c("")) #import the data and make blanks = NA

Checked the data for NA's and blanks.

summary(is.na(epl))

## Date Player.ID Player.Surname Player.Forename  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10173   
## NA's :0 NA's :0 NA's :0 TRUE :196   
## NA's :0   
## Team Team.Id Opposition Opposition.id   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Venue Position.Id Appearances Time.Played   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Starts Substitute.On Substitute.Off Goals   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## First.Goal Winning.Goal Shots.On.Target.inc.goals  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Shots.Off.Target.inc.woodwork Blocked.Shots Penalties.Taken  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Penalty.Goals Penalties.Saved Penalties.Off.Target Penalties.Not.Scored  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Direct.Free.kick.Goals Direct.Free.kick.On.Target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Direct.Free.kick.Off.Target Blocked.Direct.Free.kick  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Goals.from.Inside.Box Shots.On.from.Inside.Box Shots.Off.from.Inside.Box  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Blocked.Shots.from.Inside.Box Goals.from.Outside.Box  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Shots.On.Target.Outside.Box Shots.Off.Target.Outside.Box  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Blocked.Shots.Outside.Box Headed.Goals Headed.Shots.On.Target  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Headed.Shots.Off.Target Headed.Blocked.Shots Left.Foot.Goals  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Left.Foot.Shots.On.Target Left.Foot.Shots.Off.Target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Left.Foot.Blocked.Shots Right.Foot.Goals Right.Foot.Shots.On.Target  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Right.Foot.Shots.Off.Target Right.Foot.Blocked.Shots Other.Goals   
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Other.Shots.On.Target Other.Shots.Off.Target Other.Blocked.Shots  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Shots.Cleared.off.Line Shots.Cleared.off.Line.Inside.Area  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Shots.Cleared.off.Line.Outside.Area Goals.Open.Play Goals.from.Corners  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Goals.from.Throws Goals.from.Direct.Free.Kick Goals.from.Set.Play  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Goals.from.penalties Attempts.Open.Play.on.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Attempts.from.Corners.on.target Attempts.from.Throws.on.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Attempts.from.Direct.Free.Kick.on.target Attempts.from.Set.Play.on.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Attempts.from.Penalties.on.target Attempts.Open.Play.off.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Attempts.from.Corners.off.target Attempts.from.Throws.off.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Attempts.from.Direct.Free.Kick.off.target  
## Mode :logical   
## FALSE:10369   
## NA's :0   
##   
## Attempts.from.Set.Play.off.target Attempts.from.Penalties.off.target  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Goals.as.a.substitute Total.Successful.Passes.All  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Total.Unsuccessful.Passes.All Assists Key.Passes   
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Total.Successful.Passes.Excl.Crosses.Corners  
## Mode :logical   
## FALSE:10369   
## NA's :0   
##   
## Total.Unsuccessful.Passes.Excl.Crosses.Corners Successful.Passes.Own.Half  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Passes.Own.Half Successful.Passes.Opposition.Half  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Passes.Opposition.Half Successful.Passes.Defensive.third  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Passes.Defensive.third Successful.Passes.Middle.third  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Passes.Middle.third Successful.Passes.Final.third  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Passes.Final.third Successful.Short.Passes  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Short.Passes Successful.Long.Passes Unsuccessful.Long.Passes  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Successful.Flick.Ons Unsuccessful.Flick.Ons Successful.Crosses.Corners  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Unsuccessful.Crosses.Corners Corners.Taken.incl.short.corners  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Corners.Conceded Successful.Corners.into.Box  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Corners.into.Box Short.Corners Throw.Ins.to.Own.Player  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Throw.Ins.to.Opposition.Player Successful.Dribbles Unsuccessful.Dribbles  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Successful.Crosses.Corners.Left Unsuccessful.Crosses.Corners.Left  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Successful.Crosses.Left Unsuccessful.Crosses.Left Successful.Corners.Left  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Unsuccessful.Corners.Left Successful.Crosses.Corners.Right  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Crosses.Corners.Right Successful.Crosses.Right  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Crosses.Right Successful.Corners.Right  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.Corners.Right Successful.Long.Balls Unsuccessful.Long.Balls  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Successful.Lay.Offs Unsuccessful.Lay.Offs Through.Ball   
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Successful.Crosses.Corners.in.the.air  
## Mode :logical   
## FALSE:10369   
## NA's :0   
##   
## Unsuccessful.Crosses.Corners.in.the.air Successful.crosses.in.the.air  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.crosses.in.the.air Successful.open.play.crosses  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Unsuccessful.open.play.crosses Touches Goal.Assist.Corner  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Goal.Assist.Free.Kick Goal.Assist.Throw.In Goal.Assist.Goal.Kick  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Goal.Assist.Set.Piece Key.Corner Key.Free.Kick Key.Throw.In   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Key.Goal.Kick Key.Set.Pieces Duels.won Duels.lost   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Aerial.Duels.won Aerial.Duels.lost Ground.Duels.won Ground.Duels.lost  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Tackles.Won Tackles.Lost Last.Man.Tackle Total.Clearances  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Headed.Clearances Other.Clearances Clearances.Off.the.Line  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Blocks Interceptions Recoveries Total.Fouls.Conceded  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Fouls.Conceded.exc.handballs.pens Total.Fouls.Won  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Fouls.Won.in.Danger.Area.inc.pens Fouls.Won.not.in.danger.area  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##   
## Foul.Won.Penalty Handballs.Conceded Penalties.Conceded Offsides   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Yellow.Cards Red.Cards Goals.Conceded Goals.Conceded.Inside.Box  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Goals.Conceded.Outside.Box Saves.Made Saves.Made.from.Inside.Box  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Saves.Made.from.Outside.Box Saves.from.Penalty Catches   
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Punches Drops Crosses.not.Claimed GK.Distribution  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## GK.Successful.Distribution GK.Unsuccessful.Distribution Clean.Sheets   
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Team.Clean.sheet Error.leading.to.Goal Error.leading.to.Attempt  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Challenge.Lost Shots.On.Conceded Shots.On.Conceded.Inside.Box  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Shots.On.Conceded.Outside.Box Team.Formation Position.in.Formation  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Turnovers Dispossessed Big.Chances Big.Chances.Faced  
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Pass.Forward Pass.Backward Pass.Left Pass.Right   
## Mode :logical Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0 NA's :0   
##   
## Unsuccessful.Ball.Touch Successful.Ball.Touch Take.Ons.Overrun  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## CompId SeasId Touches.open.play.final.third  
## Mode :logical Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369 FALSE:10369   
## NA's :0 NA's :0 NA's :0   
##   
## Touches.open.play.opp.box Touches.open.play.opp.six.yards  
## Mode :logical Mode :logical   
## FALSE:10369 FALSE:10369   
## NA's :0 NA's :0   
##

The only blanks were in the forename field. Certain players, like Nani of Manchester United, go by only one name which made their forename field blank. These blanks were changed to NA's when the data was read into R.

Later on in the analysis I classified rows based on goals, but before doing that I needed to see the distribution of goals scored per game.

summary(epl$Goals) #look at the distribution of goals

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.00000 0.00000 0.09885 0.00000 4.00000

From this summary it was evident that the majority of players did not score even 1 goal in a game. It was also clear that the maximum number of goals scored by a player in 1 game was 4.

Looked at a table to see the distribution of goals scored for each of the 5 categories.

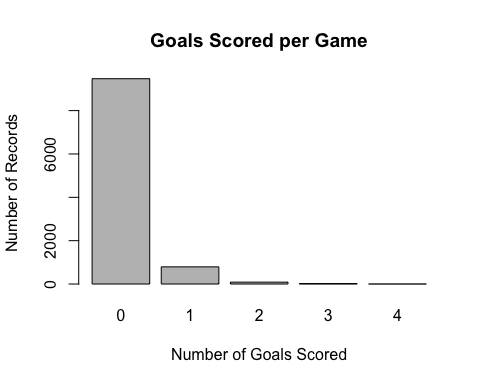
table(epl$Goals) #table of goals

##   
## 0 1 2 3 4   
## 9471 792 87 17 2

The table reaffirmed what the summary function displayed, that the majority of players did not score a goal in a game. As the number of goals per game increased the number of players who achieved this feat decreased.

A graphical representation helped illuminate the differences between the number of goals scored.

goalnumbers<- table(epl$Goals)  
barplot(goalnumbers, main = "Goals Scored per Game", xlab = "Number of Goals Scored", ylab = "Number of Records")



As expected the graph looks similar to a positively skewed distribution with a significant drop off from 0 to 1 goals scored.

A duplicate of the data was created that could be edited so that the original file did not need to be re-read into R in the event of a mistake.

epl.mod <- epl

Attached the data frame so that column names could be referenced without referencing the data frame.

attach(epl.mod)

Created a new field named Goals.New where a 1 represented that a player scored 1 or more goals that game and a 0 represented that 0 goals were scored.

epl.mod["Goals.New"] <- NA  
epl.mod$Goals.New <- ifelse(Goals >= 1, 1, 0)

Looked at a table to see the distribution of goals scored for the 2 categories.

table(epl.mod$Goals.New)

##   
## 0 1   
## 9471 898

Dropped columns that contained the same information as the dependent variable, Goals.New, such as Left Foot Goals. If these columns were not dropped the model would have had 100% accuracy because all the goal data would have been in the independent variables as well as the dependent variable.

Business understanding was used to drop other columns, e.g. Penalties Taken. Penalties highly favor the penalty taker as opposed to the goalkeeper. A team's penalty taker is selected by their manager and given the chance to take a penalty when their team is awarded a penalty kick because of a foul in the box. Players convert these penalties into goals at a very high rate. This statistic was excluded because taking a penalty results in a goal the majority of the time and goals cannot be used to predict goals.

Columns including demographic data, like Player Surname, were removed because the point of the model was to predict goals based off in game statistics not off someone's name.

epl.mod$Goals <-NULL  
epl.mod$Player.ID <- NULL  
epl.mod$Date <- NULL  
epl.mod$Player.Surname <- NULL  
epl.mod$Player.Forename <- NULL  
epl.mod$Team <- NULL  
epl.mod$Team.Id <- NULL  
epl.mod$Opposition <- NULL  
epl.mod$Opposition.id <- NULL  
epl.mod$Venue <- NULL  
epl.mod$First.Goal <- NULL  
epl.mod$Winning.Goal <- NULL  
epl.mod$Penalties.Taken <- NULL  
epl.mod$Penalty.Goals <- NULL  
epl.mod$Penalties.Saved <- NULL  
epl.mod$Penalties.Off.Target <- NULL  
epl.mod$Penalties.Not.Scored <- NULL  
epl.mod$Direct.Free.kick.Goals <- NULL  
epl.mod$Goals.from.Inside.Box <- NULL  
epl.mod$Goals.from.Outside.Box <- NULL  
epl.mod$Headed.Goals <- NULL  
epl.mod$Left.Foot.Goals <- NULL  
epl.mod$Right.Foot.Goals <- NULL  
epl.mod$Other.Goals <- NULL  
epl.mod$Goals.Open.Play <- NULL  
epl.mod$Goals.from.Corners <- NULL  
epl.mod$Goals.from.Throws <- NULL  
epl.mod$Goals.from.Direct.Free.Kick <- NULL  
epl.mod$Goals.from.Set.Play <- NULL  
epl.mod$Goals.from.penalties <- NULL  
epl.mod$Attempts.from.Penalties.on.target <- NULL  
epl.mod$Attempts.from.Penalties.off.target <- NULL  
epl.mod$Goals.as.a.substitute <- NULL

# Modeling

Loaded packages.

library(caTools) #load package for sampling  
library(e1071) #load package for svm  
library(kknn) #load package for knn  
library(pROC) #load package for roc

## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

Created train and test data frames.

set.seed(703)  
sample = sample.split(epl.mod, SplitRatio = .80)  
train.epl.mod = subset(epl.mod, sample == TRUE)  
test.epl.mod = subset(epl.mod, sample == FALSE) #create train and test data frames

Classification was performed on the Goals.New variable based off in game statistics. The two algorithms I chose to use were K Nearest Neighbor (KNN) and Support Vector Machines (SVM), both of which are ideal for classification problems.

I started with the KNN model.

knn.epl.mod <- kknn(Goals.New ~.,train.epl.mod, test.epl.mod, k= 1, kernel = "rectangular")

A comparison matrix was used to visualize the data.

table.knn <- table(pred = predict(knn.epl.mod,test.epl.mod),true = test.epl.mod$Goals.New)  
table.knn

## true  
## pred 0 1  
## 0 1856 106  
## 1 77 58

The KNN model did a great job of predicting when players scored 0 goals in a game correctly but did not do as well when predicting if they scored 1 or more goals. This could have been because the the training data set contained more 0 goal records than 1, 2, 3, and 4 goal records combined. To combat this issue I set k = 1.

The accuracy of the KNN model.

accuracy.knn <- (sum(diag(table.knn)))/sum(table.knn)  
accuracy.knn

## [1] 0.9127325

The overall KNN model had an accuracy of 91% while the 1 goal field had an accuracy of 35% and the 0 goal field had an accuracy of 96%.

The SVM model was then run. C-classification was selected since the data was not compatible with nu or one classification.

svm.epl.mod <- svm(Goals.New ~., data = train.epl.mod, type = "C-classification", kernel = "linear",  
 scale = FALSE) #svm model  
svm.output <- predict(svm.epl.mod, test.epl.mod[,-178]) #svm prediction

Used a comparison matrix to visualize the predicted vs the actual values.

table.svm <- table(pred = svm.output, true = test.epl.mod[,178]) #comparison matrix  
table.svm

## true  
## pred 0 1  
## 0 1879 79  
## 1 54 85

As in the KNN model the overall SVM model performed great when predicting if players scored 0 goals in a game correctly, 97%, but did not do as well when predicting the 1 goal field, 51% accuracy.

The accuracy of the SVM model.

accuracy.svm <- (sum(diag(table.svm)))/sum(table.svm)  
accuracy.svm

## [1] 0.9365761

The SVM model had an accuracy of 93%.

# Conclusion

The SVM model created was better suited for this data set given the higher overall accuracy percentage of the model and higher accuracy percentage of both fields.

Hope you enjoyed the read and thanks for visiting my site!