---
title: "ASIMAKÇAY"
date: "Mar 6, 2023"
format: pdf
editor: visual

Dataset

Dataset: The dataset is a CSV file named "FIFA23_official_data.csv" and can be found on Kaggle. There are rows and columns in the dataset, each of which contains different football player characteristics. Some of the variables are as follows:

ID: Player's identification number

Name: Name of the football player

Age: Age of the football player

Nationality: Nationality of the player

Overall: The overall score of the player

Potential: The player's potential score

Club: Footballer's club

Value: Transfer value of the player

Wage: Player's weekly salary

Position: Position of the player

```
{r}
install.packages("caret")
library(caret)
library(tidyverse)
getwd()
setwd("C:/Users/o.ersen/Desktop/asımaşk")
list.files()
View(data)
```

```
data <- read.csv("FIFA23_official_data.csv" ,  header = TRUE,
stringsAsFactors = FALSE)
fifa <- data
```

You can add options to executable code like this

```
{r}
fifa_clean <- fifa %>%
  select(Name, Age, Nationality, Overall, Potential, Club, Value) %>%
  mutate(ValueNumeric = parse_number(Value)) %>%
  na.omit()
```

To train the model, we will use the Overall and Potential properties as arguments and ValueNumeric as the dependent variable.

```
{r}
model <- lm(ValueNumeric ~ Overall + Potential + Age, data = fifa_clean)
library(caret)
set.seed(123)
cv <- trainControl(method = "cv", number = 10)
model_caret <- train(ValueNumeric ~ Overall + Potential + Age, data =
fifa_clean, trControl = cv, method = "lm")
summary(model_caret)
```

```
Call:
lm(formula = .outcome ~ ., data = dat)
Residuals:
   Min     1Q Median     3Q    Max
-524.2 -197.7 -105.0  168.8  807.1
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 973.7463    30.0064  32.451   <2e-16 ***
Overall     -11.1013     0.7165 -15.494   <2e-16 ***
Potential    -0.8481     0.7085  -1.197   0.2313
Age           1.9214     0.8785   2.187   0.0287 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 265.7 on 17656 degrees of freedom
Multiple R-squared: 0.09977, Adjusted R-squared: 0.09962
F-statistic: 652.2 on 3 and 17656 DF, p-value: < 2.2e-16

{r}
```
model_caret$results$Rsquared[1]
```

[1] 0.09982314

3.We will use the R-squared metric to evaluate the performance of the model. R-squared is the percentage of the independent variables explaining the variance in the dependent variable. It shows how well the model can explain.

{r}
```
# Let's calculate the R-squared value of the model
summary(model)$r.squared
```

[1] 0.09976862

4.For over- and under-fit control, we'll look at the remnants of the model. The residuals are the difference between the actual values and the model's predictions. In a good model, residuals are expected to be normally distributed and their variances to be constant.

{r}
```
# Let's calculate the residuals
resid <- residuals(model)
# Let's plot the histogram of the residues
hist(resid)
```

5.Let's create a new observation and estimate the value of the target feature using our model.

{r}

```r
#Let's create a new observation
new_obs <- data.frame(Overall = 85, Potential = 90)
#Let's do the guessing
predict(model, new_obs)
```

```
      1
4.441253
```