xgb_split

2024-12-07

```r
set.seed(42)
df_reg <- read.csv("xgb_df.csv")
#Train test split
train_indices_reg <- sample(1:nrow(df_reg),
                        0.8 * nrow(df_reg))
train_data_reg <- data.matrix(df_reg[train_indices_reg, ])
test_data_reg <- df_reg[-train_indices_reg, ]

#Define predictor and response variables in training set
train_x = data.matrix(train_data_reg
                      [,-which(colnames(train_data_reg)=="total.goals")])
train_y = train_x[,which(colnames(train_data_reg)=="total.goals")]

#Define predictor and response variables in testing set
test_x = data.matrix(test_data_reg
                      [,-which(colnames(test_data_reg)=="total.goals")])
test_y = test_x[,which(colnames(test_data_reg)=="total.goals")]

write.csv(train_x, "xgb_train_x.csv", row.names = FALSE)
write.csv(train_y, "xgb_train_y.csv", row.names = FALSE)
write.csv(test_x, "xgb_test_x.csv", row.names = FALSE)
write.csv(test_y, "xgb_test_y.csv", row.names = FALSE)
write.csv(train_data_reg, "xgb_train.csv", row.names = FALSE)
write.csv(test_data_reg, "xgb_test.csv", row.names = FALSE)
```