Bland Altman Analysis

Magali Eisik

Contents

```r
#Load required packages
library(knitr) # for dynamic reports
library(kableExtra) # for style tables
library(ggplot2) # Data visualization
library(blandr) #Bland-Altman plots
```

0.1   Performance Metrics

```r
metrics_models <- data.frame(
Model = c("Elastic Net", "Random Forest"),
R2 = c(0.82, 0.60),
MAE = c(3.50, 5.03),
RMSE = c(4.46, 6.19)
)

#Display table

metrics_models %>%
kbl(
caption = "Performance Metrics for Machine Learning Models",
align = "c",
booktabs = TRUE
) %>%
kable_styling(
full_width = FALSE,
position = "center",
latex_options = c("striped", "hold_position")
)
```

0.2   Bland-Altman Analysis

Table 1: Performance Metrics for Machine Learning Models

Model	R2	MAE	RMSE
Elastic Net	0.82	3.50	4.46
Random Forest	0.60	5.03	6.19

```r
#Load datasets with predicted ages
#Predicted ages from Random forest
rf_pred <- read.csv("RF_Predicted_Ages.csv", header = TRUE)
#Predicted ages from Elastic Net
enet_pred <- read.csv("ENet_Predicted_Ages.csv", header = TRUE)

#Verify dimensions and first rows
dim(rf_pred)
```

```
## [1] 2664    3
```

```r
dim(enet_pred)
```

```
## [1] 2664    3
```

```r
head(rf_pred,3)
```

```
##     SampleID Observed_Age  Pred_RF
## 1 GSM1343050         57.9 54.65600
## 2 GSM1343051         42.0 44.99833
## 3 GSM1343052         47.4 47.48900
```

```r
head(enet_pred,3)
```

```
##     SampleID Observed_Age Pred_ENet
## 1 GSM1343050         57.9  56.70219
## 2 GSM1343051         42.0  45.01621
## 3 GSM1343052         47.4  47.74842
```

```r
#Combine predictions into one data frame
pred_comparison <- merge(
  rf_pred[, c("SampleID", "Observed_Age", "Pred_RF")],
  enet_pred[, c("SampleID", "Pred_ENet")],
  by = "SampleID"
)

pred_table <- pred_comparison[, c("Observed_Age", "Pred_RF", "Pred_ENet")]

# Display first rows
head(pred_table,3)
```

```
##    Observed_Age  Pred_RF Pred_ENet
## 1        57.9 54.65600  56.70219
## 2        42.0 44.99833  45.01621
## 3        47.4 47.48900  47.74842
```

```r
#Bland-Altman Plot: Random Forest vs Elastic Net

p <- blandr.draw(
pred_table$Pred_RF,
pred_table$Pred_ENet,
method1name = "Random Forest",
method2name = "Elastic Net",
plotTitle = "Bland-Altman Plot: RF vs ENet Age Predictions"
)


p +
theme_minimal() +
theme(
plot.title = element_text(hjust = 0.5, size=14, face="bold")
) +
geom_point(color="steelblue", size=2) +
geom_hline(yintercept=0, linetype="dashed", color="coral1")
```

Bland–Altman Plot: RF vs ENet Age Predictions

3
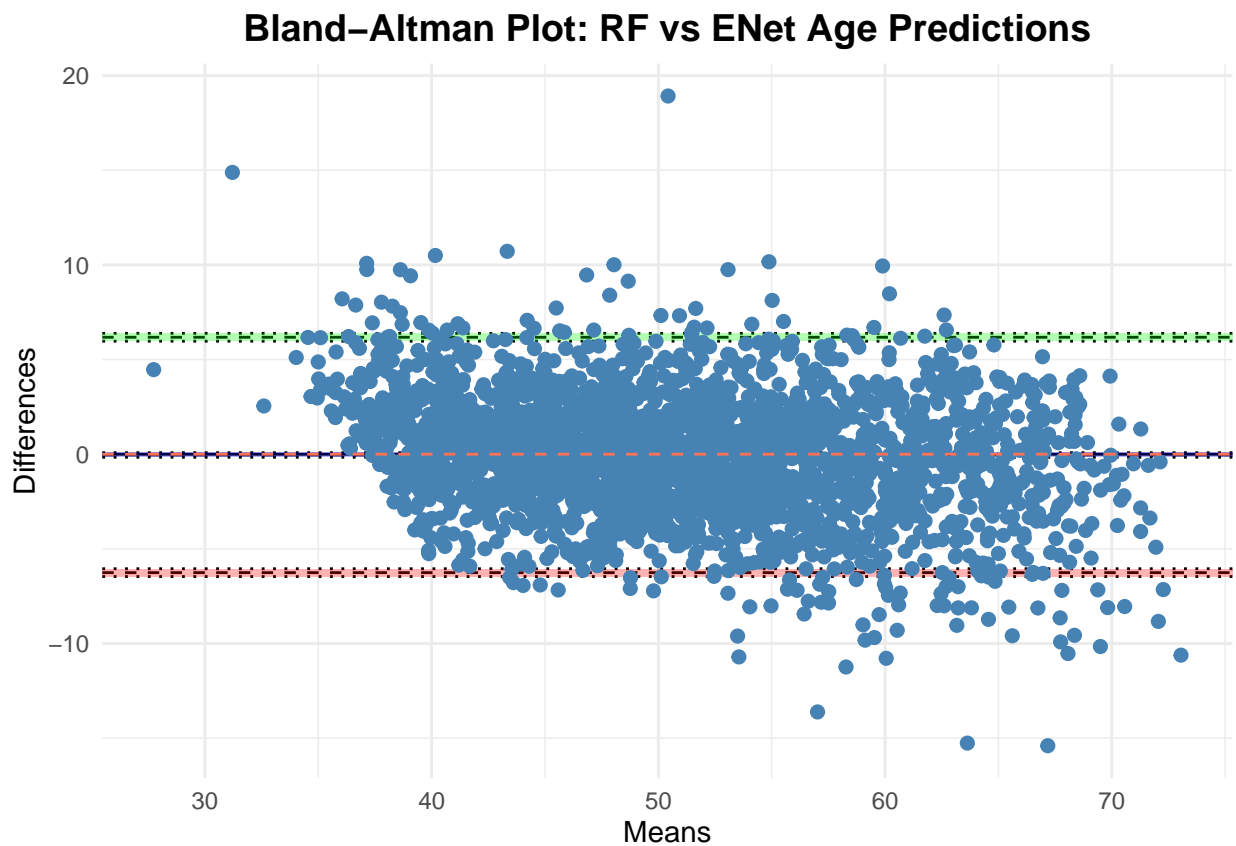
```r
# Bland-Altman statistics

ba_stats <- blandr.statistics(
pred_comparison$Pred_RF,
pred_comparison$Pred_ENet
)

# Print  summary
print(ba_stats)
```

```
## Bland-Altman Statistics
## =====================
## t = -0.62995, df = 2663, p-value = 0.5288
## alternative hypothesis: true bias is not equal to 0
##
## =====================
## Number of comparisons:  2664
## Maximum value for average measures:  73.0536
## Minimum value for average measures:  27.73673
## Maximum value for difference in measures:  18.92182
## Minimum value for difference in measures:  -15.39785
##
## Bias:  -0.03871381
## Standard deviation of bias:  3.171953
##
## Standard error of bias:  0.06145535
## Standard error for limits of agreement:  0.105041
##
## Bias:  -0.03871381
## Bias- upper 95% CI:  0.08179124
## Bias- lower 95% CI:  -0.1592188
##
## Upper limit of agreement:  6.178315
## Upper LOA- upper 95% CI:  6.384285
## Upper LOA- lower 95% CI:  5.972345
##
## Lower limit of agreement:  -6.255743
## Lower LOA- upper 95% CI:  -6.049772
## Lower LOA- lower 95% CI:  -6.461713
##
## =====================
## Derived measures:
## Mean of differences/means:  0.1770199
## Point estimate of bias as proportion of lowest average:  -0.139576
## Point estimate of bias as proportion of highest average -0.0529937
## Spread of data between lower and upper LoAs:  12.43406
## Bias as proportion of LoA spread:  -0.311353
##
## =====================
## Bias:
##  -0.03871381  ( -0.1592188  to  0.08179124 )
## ULoA:
##  6.178315  ( 5.972345  to  6.384285 )
```

```
## LLoA:
##  -6.255743  ( -6.461713  to  -6.049772 )
```

Random Forest and Elastic Net models were evaluated for predicting age. The Bland–Altman plot indicated that differences between predicted and observed ages were small and evenly distributed around zero, with a mean difference of –0.04 years and 95% limits of agreement from –6.26 to 6.18 years . A t-test confirmed that the bias was not significant (bias = –0.039, p = 0.529). While both models showed good agreement across the age range, Elastic Net demonstrated superior performance across all evaluated metrics, achieving higher accuracy (R^2 = 0.82 vs. 0.60) and lower errors (MAE = 3.50 vs. 5.03 years, RMSE = 4.46 vs. 6.19 years) compared to Random Forest

0.3 References

Bland, J. M. (2020). blandr: Bland-Altman Method Comparison Analyses. R package version 1.0.2. Available at: https://cran.r-project.org/package=blandr