

Ensemble Project

2024-03-29

```
# Set CRAN mirror
options(repos = c(CRAN = "https://cran.r-project.org"))

# Install and load necessary packages
install.packages("ggplot2")

## Installing package into 'C:/Users/USER/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\USER\AppData\Local\Temp\Rtmp0Unoc8\downloaded_packages

library(ggplot2)
```

Performance Comparison of RF on MI-based and SelectKBest with ANOVA F-value Scoring Selection Method

```
# Define the performance accuracy values
mi_based_accuracy <- c(0.99986, 0.99972, 0.99487, 0.99488)
selectkbest_accuracy <- c(0.96772, 0.9677, 0.94033, 0.9374)

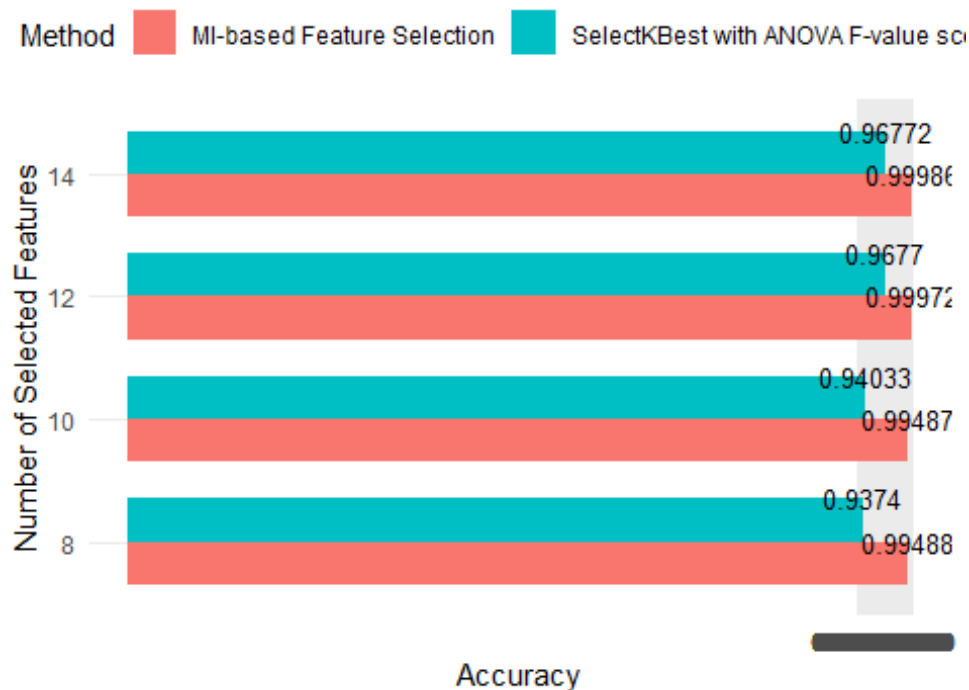
# Define the number of selected features
num_features <- c(14, 12, 10, 8)

# Create a data frame
data <- data.frame(Method = rep(c("MI-based Feature Selection", "SelectKBest
with ANOVA F-value scoring"), each = 4),
                  Num_Features = rep(num_features, 2),
                  Accuracy = c(mi_based_accuracy, selectkbest_accuracy))

# Plotting the multiple bar chart
ggplot(data, aes(x = factor(Num_Features), y = Accuracy, fill = Method, label
= round(Accuracy, 5))) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +
  geom_text(position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5)
+
  labs(x = "Number of Selected Features", y = "Accuracy",
       title = "Performance Comparison of Random Forest with Different Featur
```

```
e Selection Methods") +
  scale_y_continuous(breaks = seq(0.93000, 1.00000, by = 0.00050), labels = scales::number_format(accuracy = 0.00001)) +
  coord_flip() +
  theme_minimal() +
  theme(legend.position = "top")
```

Performance Comparison of Random Forest with Differ



Performance Comparison of XGBoost on MI-based and SelectKBest with ANOVA F-value Scoring Selection Method

```
# Define the performance accuracy values
mi_based_accuracy <- c(0.99955, 0.99898, 0.9945, 0.9945)
selectkbest_accuracy <- c(0.98831, 0.98769, 0.97116, 0.96604)

# Define the number of selected features
num_features <- c(14, 12, 10, 8)

# Create a data frame
data <- data.frame(Method = rep(c("MI-based Feature Selection", "SelectKBest
with ANOVA F-value scoring"), each = 4),
  Num_Features = rep(num_features, 2),
  Accuracy = c(mi_based_accuracy, selectkbest_accuracy))

# Plotting the multiple bar chart
ggplot(data, aes(x = factor(Num_Features), y = Accuracy, fill = Method, label
```



```

Accuracy = c(mi_based_accuracy, selectkbest_accuracy))

# Plotting the multiple bar chart
ggplot(data, aes(x = factor(Num_Features), y = Accuracy, fill = Method, label = round(Accuracy, 5))) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +
  geom_text(position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5)
+
  labs(x = "Number of Selected Features", y = "Accuracy",
       title = "Performance Comparison of NB with Different Feature Selection
Methods") +
  scale_y_continuous(breaks = seq(0.93000, 1.00000, by = 0.00050), labels = scales::number_format(accuracy = 0.00001)) +
  coord_flip() +
  theme_minimal() +
  theme(legend.position = "top")

```

