Week1 Assignment KNN Time Series

Mohit Manjaria

9/24/2020

ALY 6020 Predictive Analytics

Week1 Assignment KNN Time Series

Mohit Manjaria

Instructor: Marco Montes de Oca

Winter 2021

January 28th 2021

Northeastern University

Introduction

In this assignment, a dataset of search interest of all categories for Predictive analytics term from January 2061 to January 2021 is selected for K-Nearest Neighbors Analysis. In this dataset, the numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Analysis

In this part, time series forecasting with KNN regression will be performed. According to the auto-regressive model and model requirements that are given, three dimension, four dimension, five dimension and six dimension models with K values from 1 to 10 will be explored.

Step 1: Installing libraries

```
library(tinytex)
library(neighbr)
library(readr)
```

```r
library(tsfknn)
library(zoo)

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library(ggplot2)
```

Step 2: Importing the data and checking the data

```r
knnt <- read.csv("C:/Users/mohit/Downloads/multiTimeline.csv")

head(knnt)

##        Week Predictive.analytics...United.States.
## 1 1/31/2016                                    60
## 2  2/7/2016                                    54
## 3 2/14/2016                                    58
## 4 2/21/2016                                    53
## 5 2/28/2016                                    43
## 6  3/6/2016                                    54

dim(knnt)

## [1] 259   2

summary(knnt)

##      Week            Predictive.analytics...United.States.
##  Length:259         Min.   :  9.00
##  Class :character   1st Qu.: 42.00
##  Mode  :character   Median : 52.00
##                     Mean   : 53.07
##                     3rd Qu.: 63.00
##                     Max.   :100.00
```

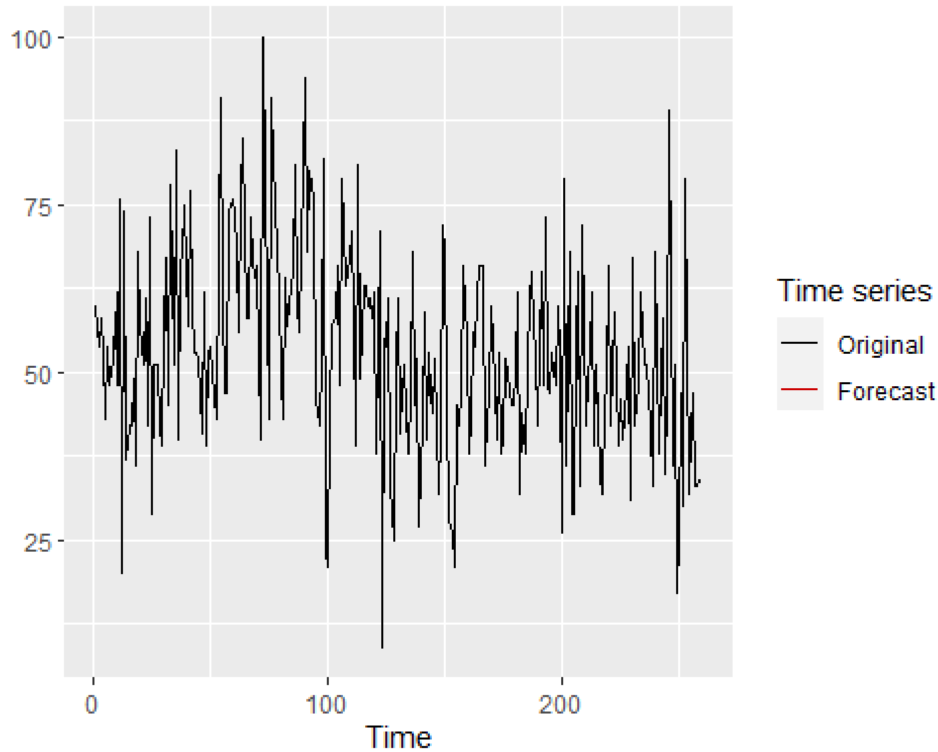
```r
#KNN Model
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:2, k=1)
pred$prediction

## Time Series:
## Start = 260
## End = 260
```

```
## Frequency = 1
## [1] 34
```

```r
pred$neighbors
```

```
## [1] 259
```

```r
#Plotting time series
autoplot(pred, h=1)
```




```r
#Calculating accuracy for k = 1
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE
## 11.00000 11.00000 32.35294
```

```r
ro$predictions
```

```
##      h=1
## [1,]  45
```

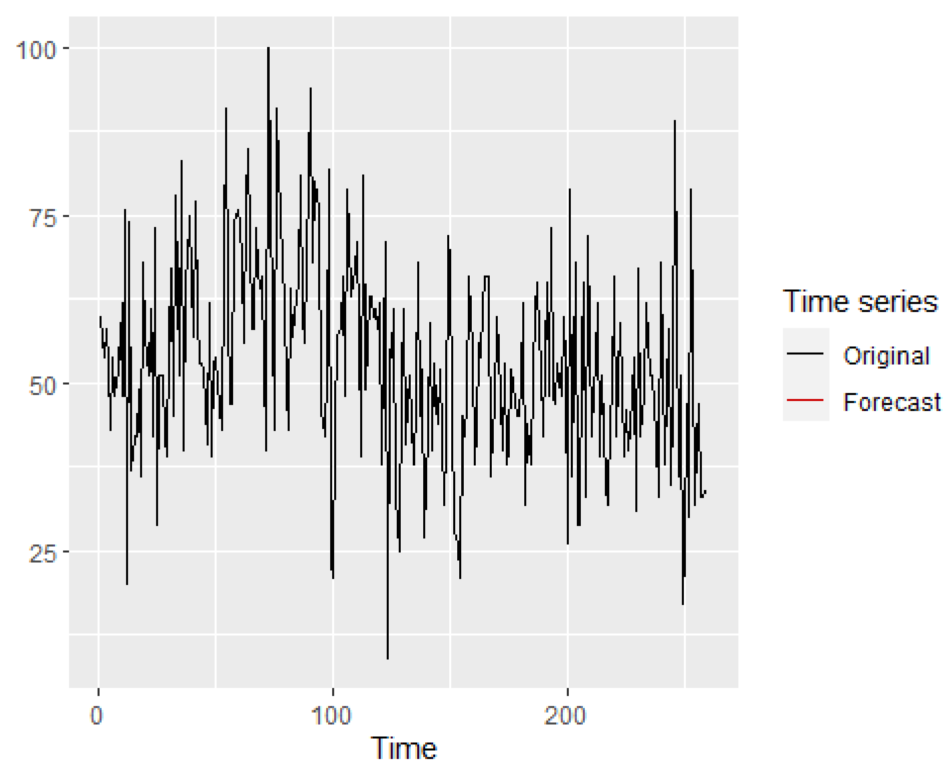
```r
ro$h_accu
```

```
##           h=1
## RMSE 11.00000
## MAE  11.00000
## MAPE 32.35294
```

#Calculating Euclidean Distance install.packages(“philentropy”) library(philentropy)

#knn.dist(knnt, dist.meth = “euclidean”, p = 2)

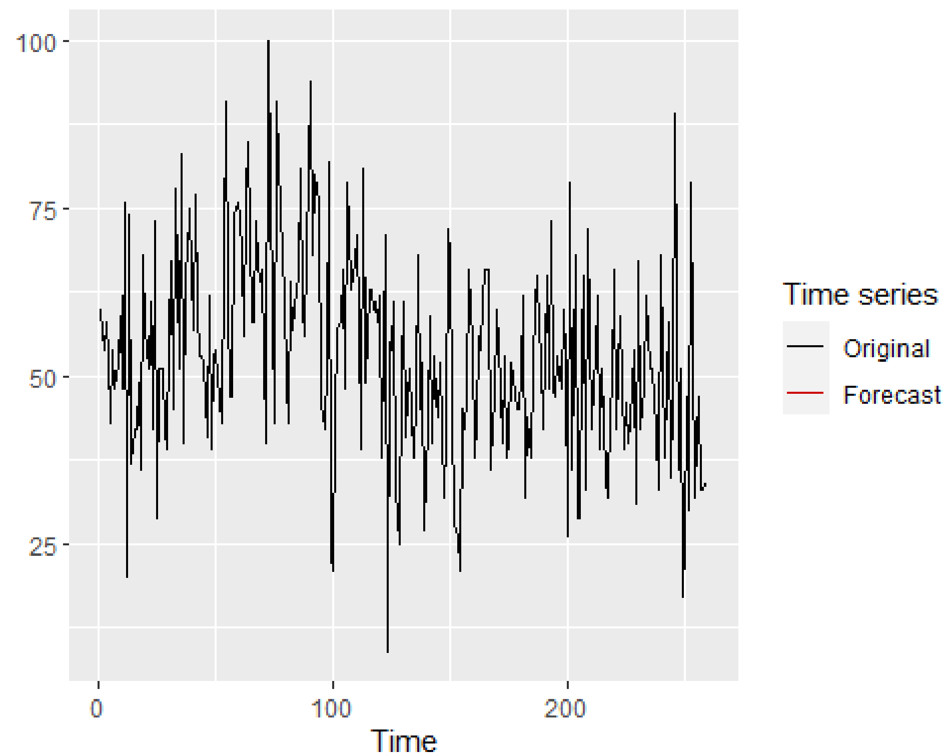
```
#For n=2 or K = 3
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:2, k=3)
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 44.66667

pred$neighbors

## [1] 259 218 206

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE
## 16.33333 16.33333 48.03922

ro$predictions

##           h=1
## [1,] 50.33333

ro$h_accu

##           h=1
## RMSE 16.33333
## MAE  16.33333
## MAPE 48.03922
```

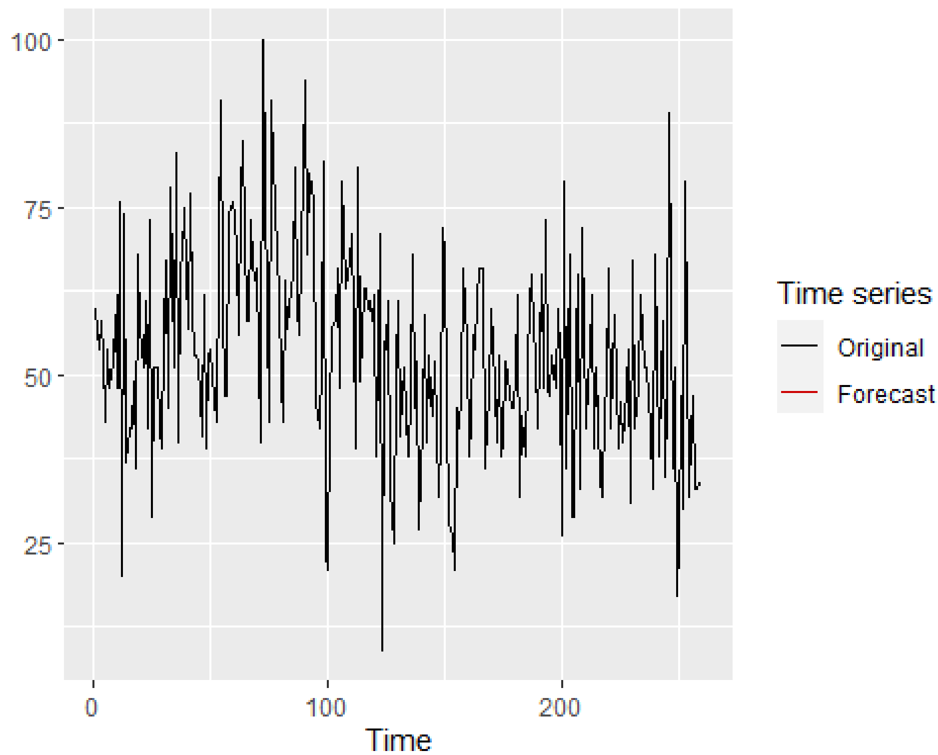
```
#For n=2, k = 5

pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:2, k=5)
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 50.6

pred$neighbors

## [1] 259 218 206 149 256

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 20.00000 20.00000 58.82353

ro$predictions

##      h=1
## [1,]  54

ro$h_accu

##           h=1
## RMSE 20.00000
## MAE  20.00000
## MAPE 58.82353
```

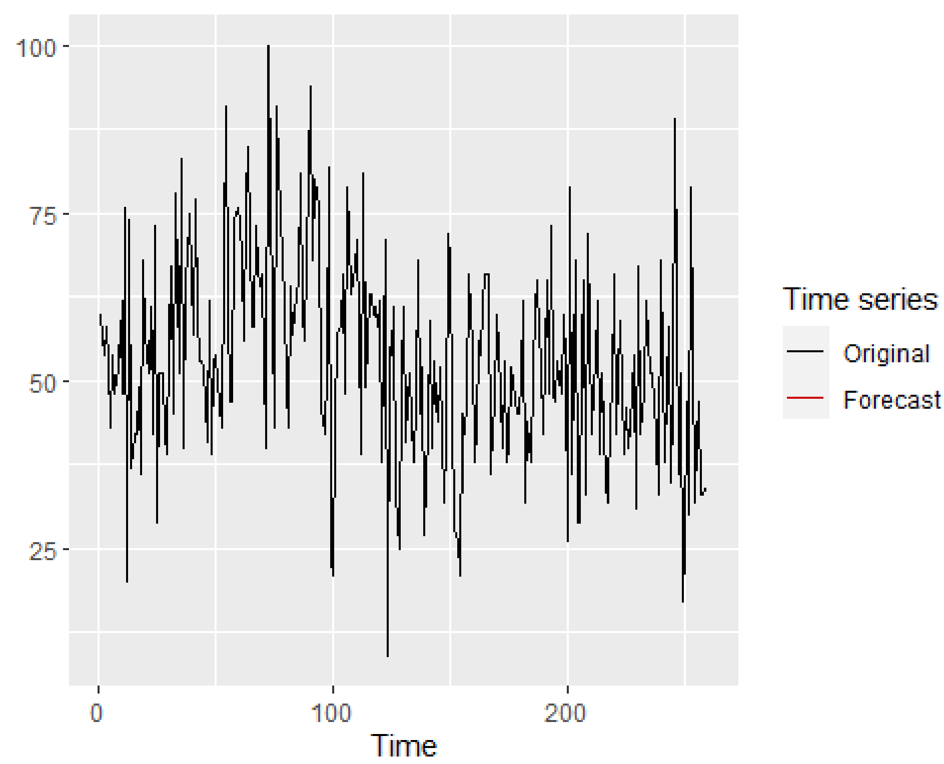
```
#for n = 2, k = 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:2, k=7)
pred$prediction

## Time Series:
## Start = 260
```

```
## End = 260
## Frequency = 1
## [1] 49.57143

pred$neighbors

## [1] 259 218 206 149 256 176  16

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 15.00000 15.00000 44.11765

ro$predictions

##      h=1
## [1,]  49

ro$h_accu

##           h=1
## RMSE 15.00000
## MAE  15.00000
## MAPE 44.11765
```

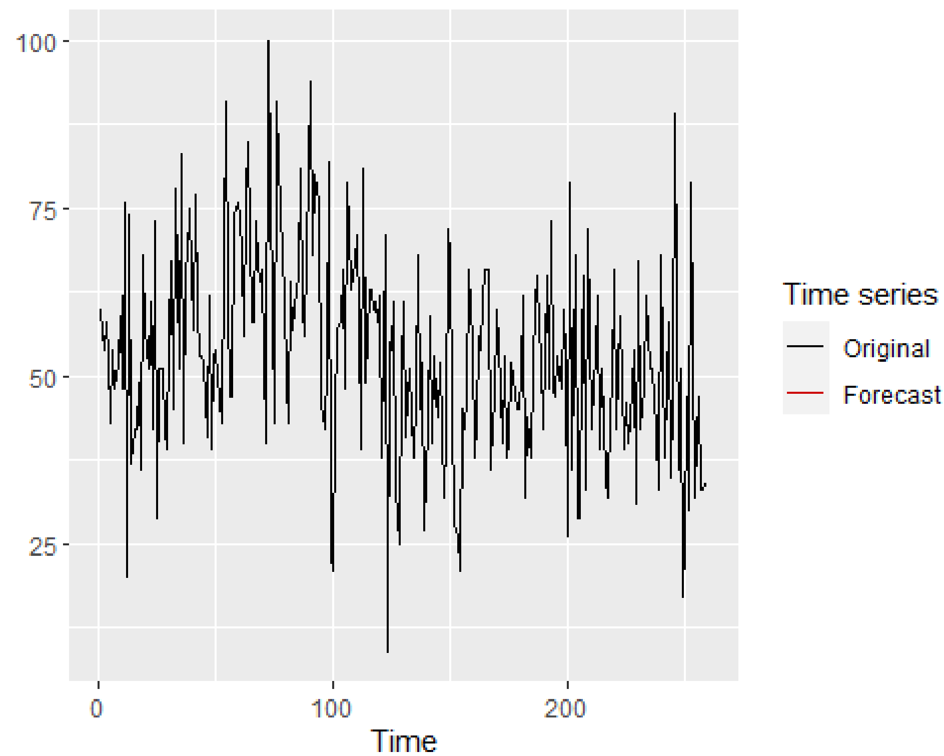
```r
#for n = 3, k = 1

pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:3, k=1)
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 48

pred$neighbors

## [1] 219

#Plotting time series
autoplot(pred, h=1)
```




```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 11.00000 11.00000 32.35294
```

```r
ro$predictions

##      h=1
## [1,]  45

ro$h_accu

##           h=1
## RMSE 11.00000
## MAE  11.00000
## MAPE 32.35294
```

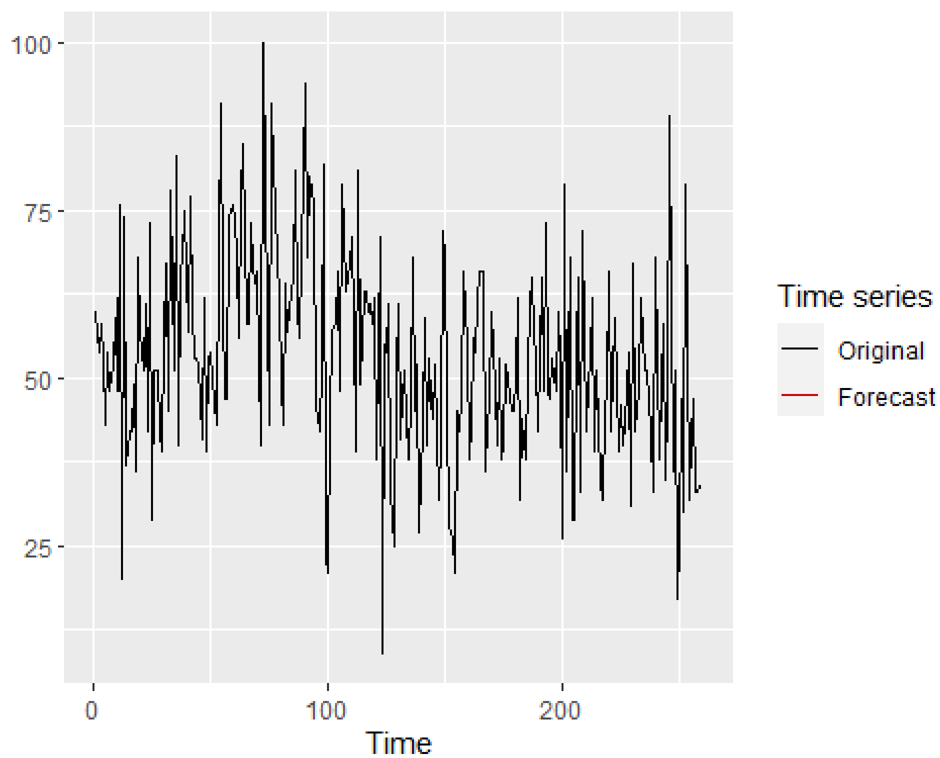
```r
#for n = 3, k = 3
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:3, k=3)
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 49.33333

pred$neighbors

## [1] 219 185  17

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 10.66667 10.66667 31.37255

ro$predictions

##           h=1
## [1,] 44.66667

ro$h_accu

##           h=1
## RMSE 10.66667
## MAE  10.66667
## MAPE 31.37255
```

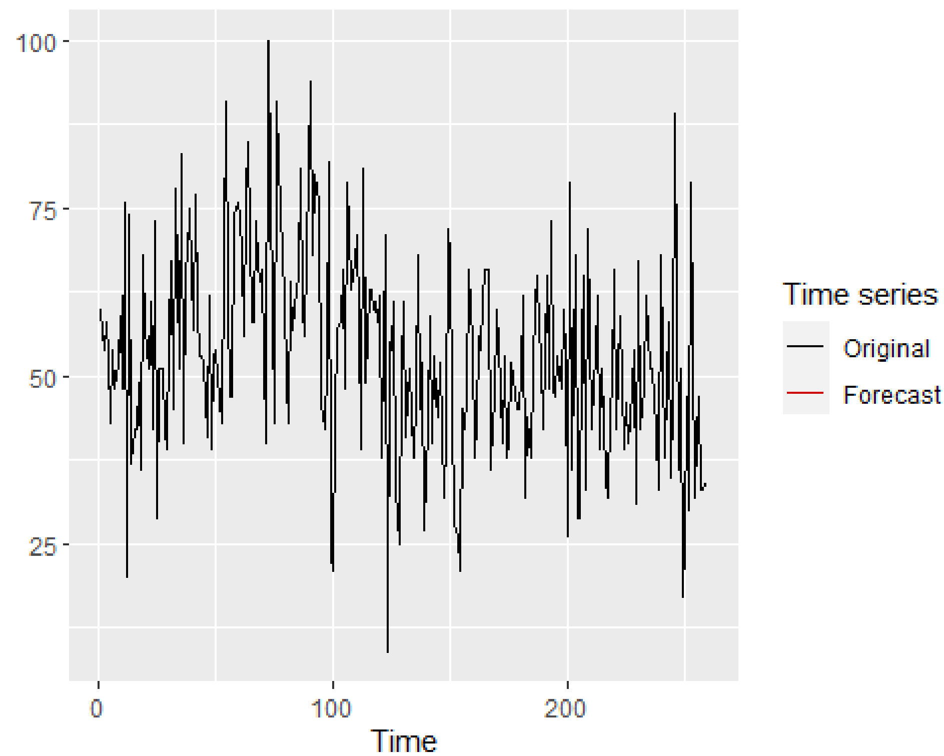
```
# for n =3, k = 5
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:3, k=5)
pred$prediction

## Time Series:
## Start = 260
```

```
## End = 260
## Frequency = 1
## [1] 43

pred$neighbors

## [1] 219 185  17 259 257

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 17.60000 17.60000 51.76471

ro$predictions

##       h=1
## [1,] 51.6

ro$h_accu

##           h=1
## RMSE 17.60000
## MAE  17.60000
## MAPE 51.76471
```

```r
#For n = 3, K = 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
= 1:3, k=7)
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 43.28571

pred$neighbors

## [1] 219 185  17 259 257 227 155

nearest_neighbors(pred)

## $instance
## Lag 3 Lag 2 Lag 1
##    33    33    34
##
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
## 4    47    33    33 34
## 5    32    41    47 33
## 6    39    46    40 43
## 7    28    26    21 45

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 18.57143 18.57143 54.62185

ro$predictions

##          h=1
## [1,] 52.57143

ro$h_accu

##          h=1
## RMSE 18.57143
## MAE  18.57143
## MAPE 54.62185
```

```
#for n=2, k = 1, 3
#Combining several models with different k parameters
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(1,3))
pred$prediction
```

```
## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 39.33333
```

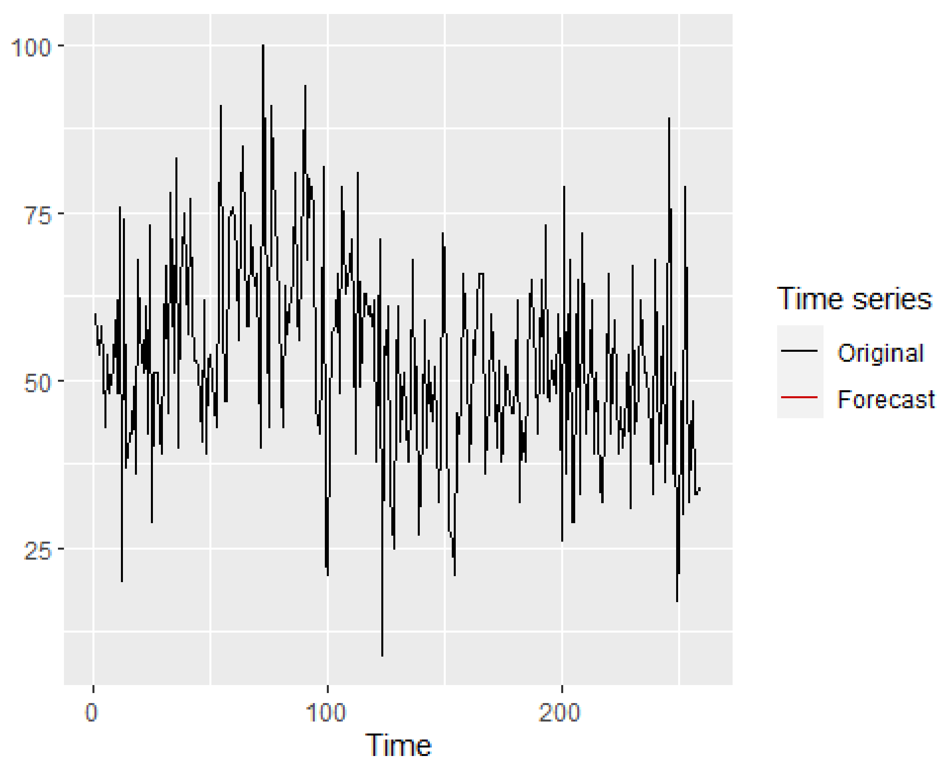
```
pred$neighbors
```

```
## [1] 259 218 206
```

```
nearest_neighbors(pred)
```

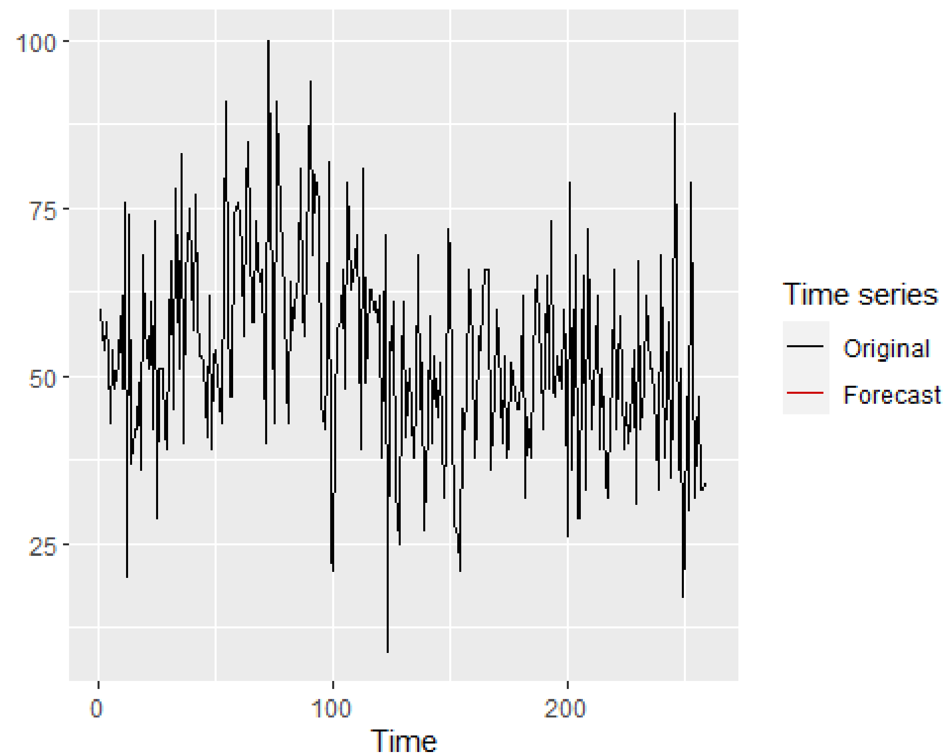
```
## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
```

```
#Plotting time series
autoplot(pred, h=1)
```



```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 13.66667 13.66667 40.19608

ro$predictions

##           h=1
## [1,] 47.66667

ro$h_accu

##           h=1
## RMSE 13.66667
## MAE  13.66667
## MAPE 40.19608
```

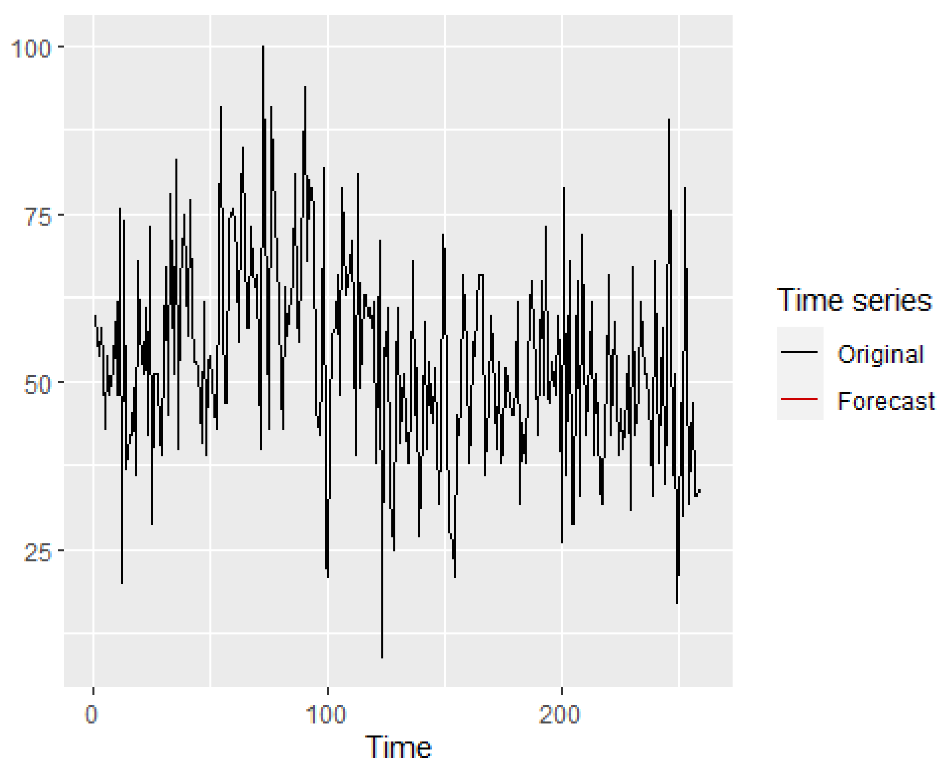
```r
#for n=2, k = 1, 5
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(1,5))
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 42.3

pred$neighbors

## [1] 259 218 206 149 256

nearest_neighbors(pred)

## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 15.50000 15.50000 45.58824

ro$predictions

##       h=1
## [1,] 49.5

ro$h_accu

##           h=1
## RMSE 15.50000
## MAE  15.50000
## MAPE 45.58824
```

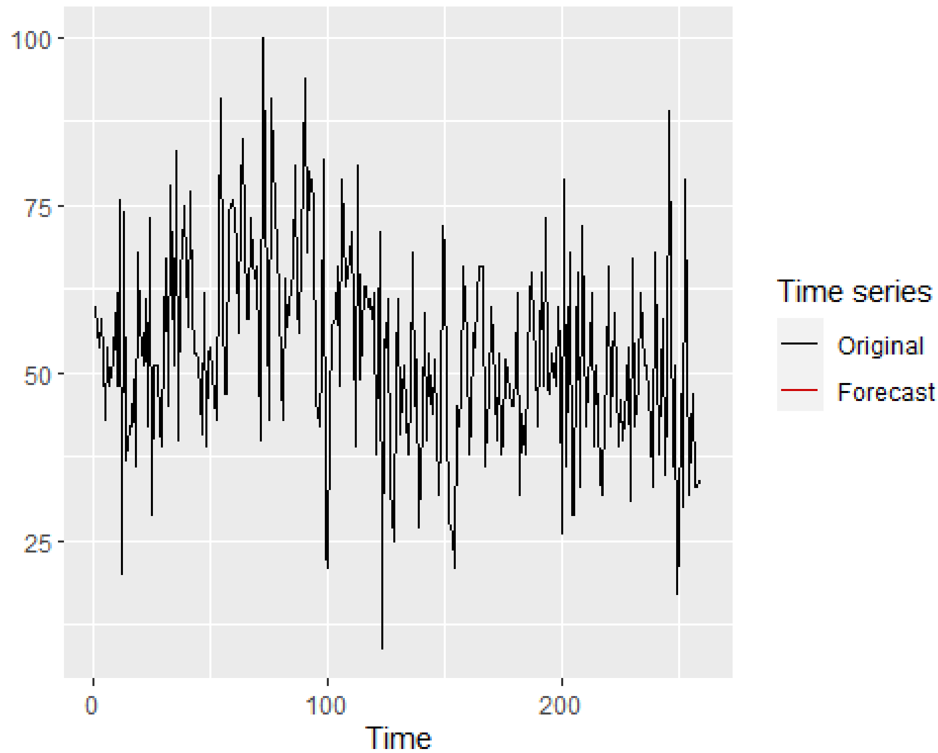
```
#for n=2, k = 1, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(1,7))
pred$prediction

## Time Series:
## Start = 260
```

```
## End = 260
## Frequency = 1
## [1] 41.78571

pred$neighbors

## [1] 259 218 206 149 256 176  16

nearest_neighbors(pred)

## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47
## 6    38    40 52
## 7    37    42 42

#Plotting time series
autoplot(pred, h=1)
```



```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 13.00000 13.00000 38.23529

ro$predictions

##      h=1
## [1,]  47

ro$h_accu

##           h=1
## RMSE 13.00000
## MAE  13.00000
## MAPE 38.23529
```

```r
#for n = 2, k = 3, 5
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(3,5))
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 47.63333

pred$neighbors

## [1] 259 218 206 149 256

nearest_neighbors(pred)

## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE
## 18.16667 18.16667 53.43137
```

```
ro$predictions
```

```
##            h=1
## [1,] 52.16667
```

```
ro$h_accu
```

```
##            h=1
## RMSE 18.16667
## MAE  18.16667
## MAPE 53.43137
```

```
#for n = 2, k = 3, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(3,7))
pred$prediction
```

```
## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 47.11905
```

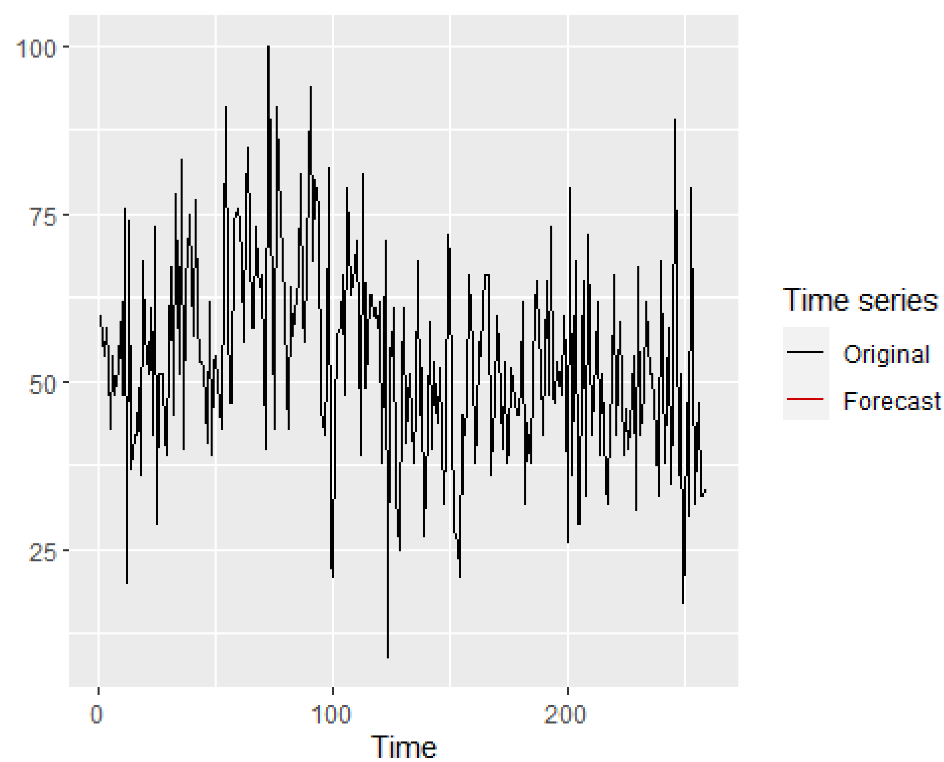
```
pred$neighbors
```

```
## [1] 259 218 206 149 256 176  16
```

```
nearest_neighbors(pred)
```

```
## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47
## 6    38    40 52
## 7    37    42 42
```

```
#Plotting time series
autoplot(pred, h=1)
```



```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE
## 15.66667 15.66667 46.07843
```

```r
ro$predictions
```

```
##           h=1
## [1,] 49.66667
```

```r
ro$h_accu
```

```
##           h=1
## RMSE 15.66667
## MAE  15.66667
## MAPE 46.07843
```

```r
#for n = 2, k = 5, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(5,7))
pred$prediction
```

```
## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 50.08571
```

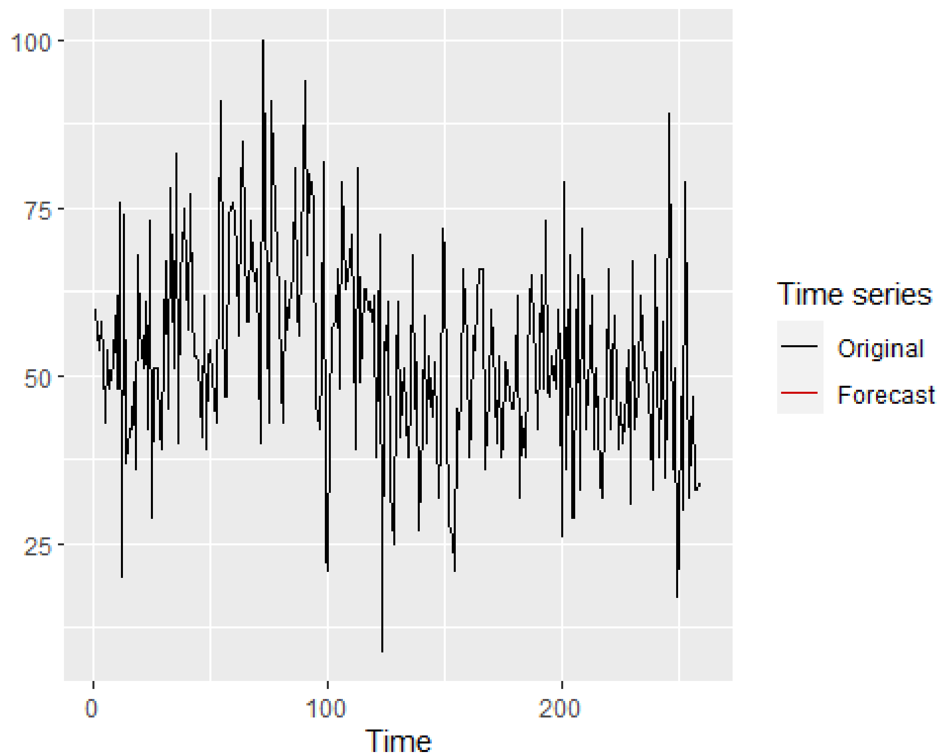
```r
pred$neighbors
```

```
## [1] 259 218 206 149 256 176  16
```

```r
nearest_neighbors(pred)
```

```
## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47
## 6    38    40 52
## 7    37    42 42
```

```
#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 17.50000 17.50000 51.47059

ro$predictions

##       h=1
## [1,] 51.5

ro$h_accu

##           h=1
## RMSE 17.50000
## MAE  17.50000
## MAPE 51.47059
```

```
#for n = 3, k = 1, 3
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:3, k =c(1,3))
pred$prediction
```

```
## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 48.66667
```

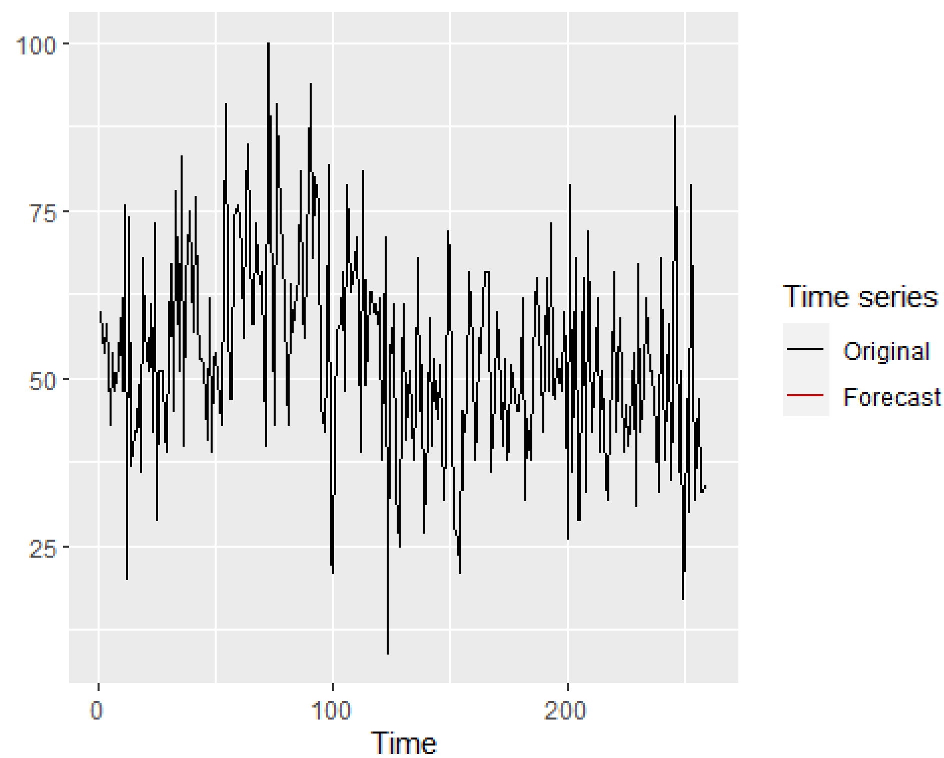
```
pred$neighbors
```

```
## [1] 219 185  17
```

```
nearest_neighbors(pred)
```

```
## $instance
## Lag 3 Lag 2 Lag 1
##    33    33    34
##
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
```

```
#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE
## 10.83333 10.83333 31.86275
```

```r
ro$predictions
```

```
##           h=1
## [1,] 44.83333
```

```r
ro$h_accu
```

```
##           h=1
## RMSE 10.83333
## MAE  10.83333
## MAPE 31.86275
```

```r
#for n=3, k = 1, 5
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:3, k =c(1,5))
pred$prediction
```

```
## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 45.5
```

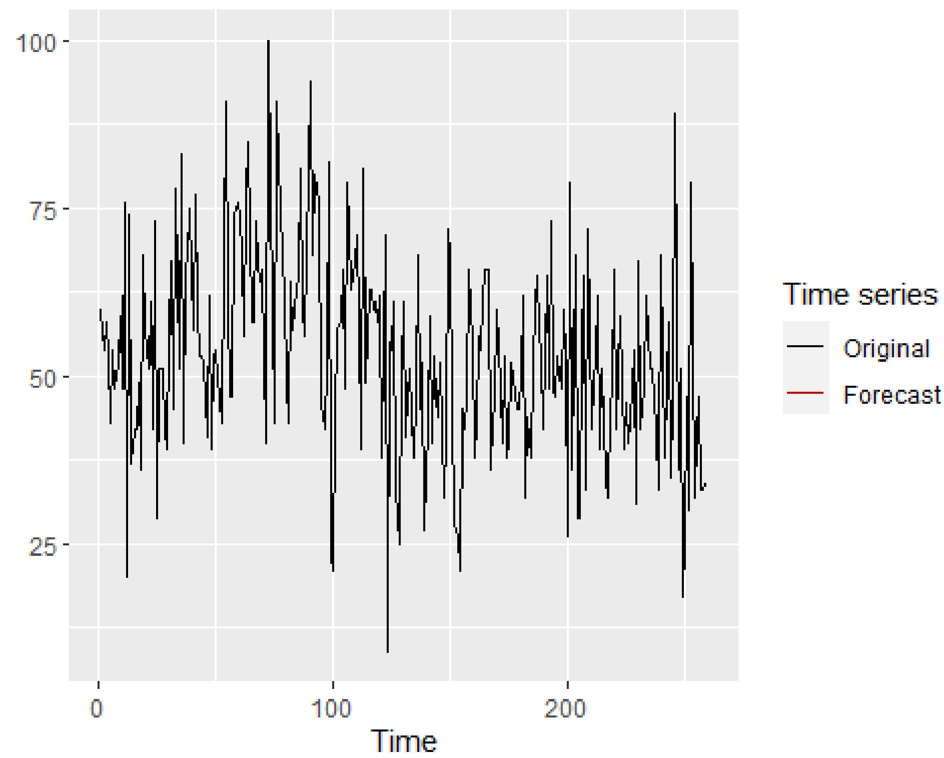
```r
pred$neighbors
```

```
## [1] 219 185  17 259 257
```

```r
nearest_neighbors(pred)
```

```
## $instance
## Lag 3 Lag 2 Lag 1
##    33    33    34
##
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
## 4    47    33    33 34
## 5    32    41    47 33
```

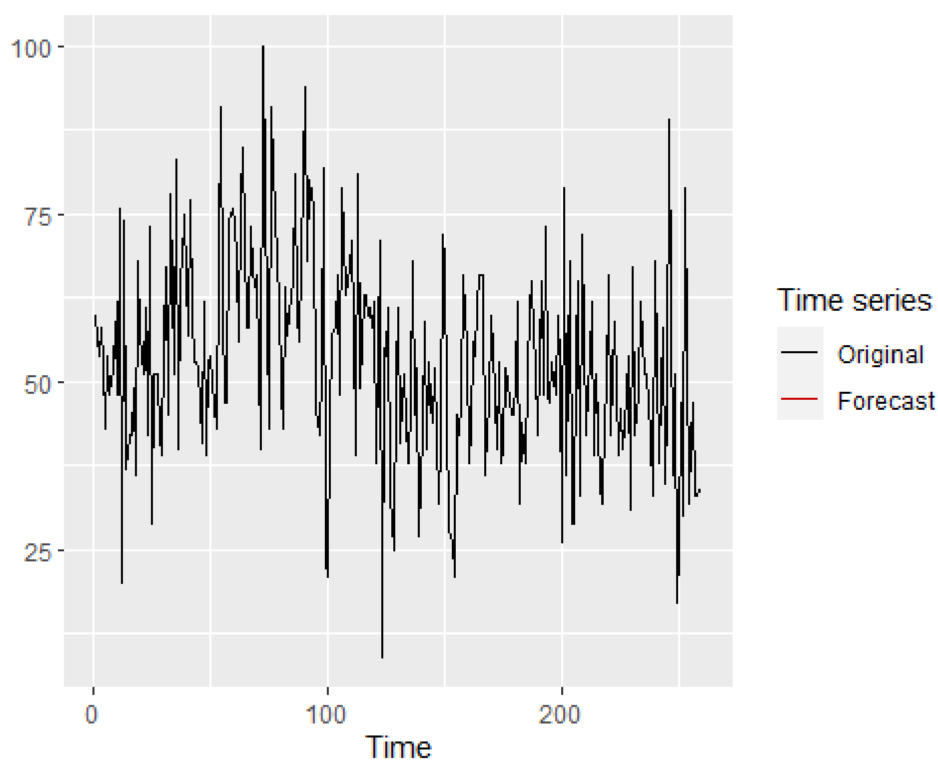
```r
#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 14.30000 14.30000 42.05882

ro$predictions

##       h=1
## [1,] 48.3

ro$h_accu

##           h=1
## RMSE 14.30000
## MAE  14.30000
## MAPE 42.05882
```

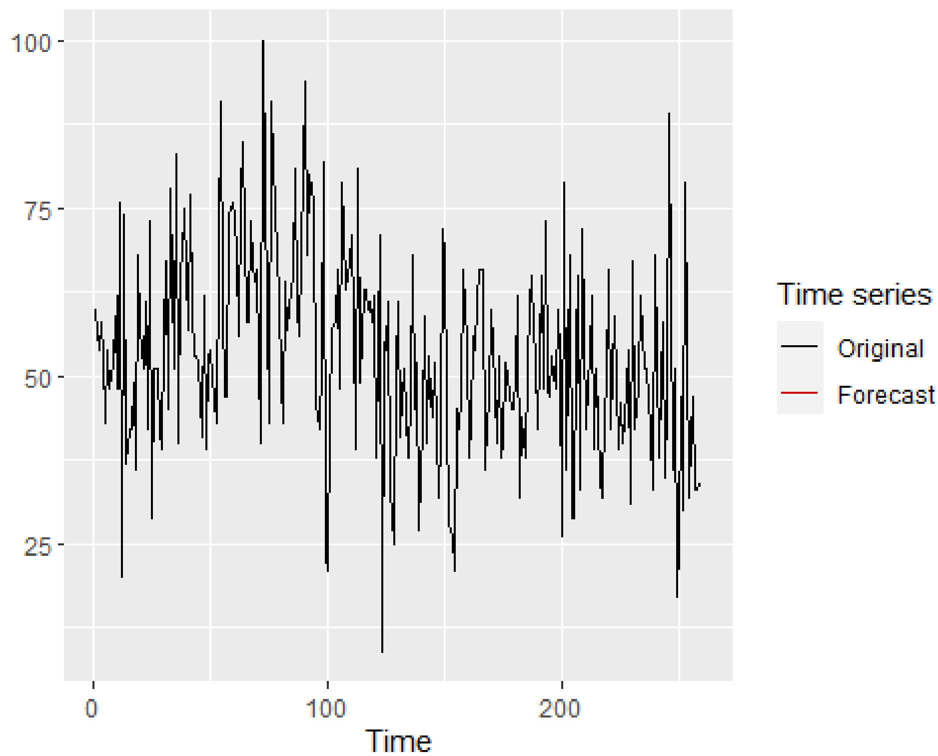
```
#for n=3, k = 1, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:3, k =c(1,7))
pred$prediction

## Time Series:
## Start = 260
```

```
## End = 260
## Frequency = 1
## [1] 45.64286

pred$neighbors

## [1] 219 185  17 259 257 227 155

nearest_neighbors(pred)

## $instance
## Lag 3 Lag 2 Lag 1
##    33    33    34
##
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
## 4    47    33    33 34
## 5    32    41    47 33
## 6    39    46    40 43
## 7    28    26    21 45

#Plotting time series
autoplot(pred, h=1)
```



```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 14.78571 14.78571 43.48739

ro$predictions

##            h=1
## [1,] 48.78571

ro$h_accu

##           h=1
## RMSE 14.78571
## MAE  14.78571
## MAPE 43.48739
```

```r
#for n = 3, k = 3, 5
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:3, k =c(3,5))
pred$prediction

## Time Series:
## Start = 260
## End = 260
## Frequency = 1
## [1] 46.16667

pred$neighbors

## [1] 219 185  17 259 257

nearest_neighbors(pred)

## $instance
## Lag 3 Lag 2 Lag 1
##    33    33    34
##
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
## 4    47    33    33 34
## 5    32    41    47 33

#Plotting time series
autoplot(pred, h=1)
```




```
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE
## 14.13333 14.13333 41.56863

ro$predictions

##           h=1
## [1,] 48.13333

ro$h_accu

##           h=1
## RMSE 14.13333
## MAE  14.13333
## MAPE 41.56863
```

```
#for n = 3, k = 3, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:2, k =c(3,7))
pred$prediction

## Time Series:
## Start = 260
```

```
## End = 260
## Frequency = 1
## [1] 47.11905
```

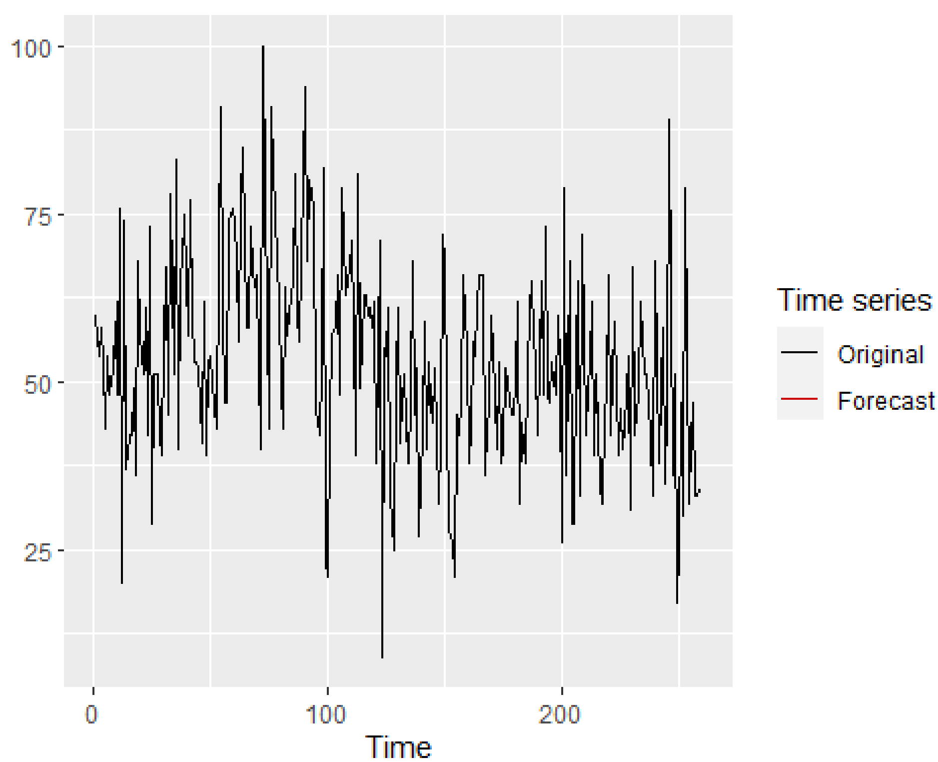
```r
pred$neighbors
```

```
## [1] 259 218 206 149 256 176  16
```

```r
nearest_neighbors(pred)
```

```
## $instance
## Lag 2 Lag 1
##    33    34
##
## $nneighbors
##   Lag 2 Lag 1 H1
## 1    33    33 34
## 2    35    32 45
## 3    29    29 55
## 4    32    41 72
## 5    32    41 47
## 6    38    40 52
## 7    37    42 42
```

```r
#Plotting time series
autoplot(pred, h=1)
```



```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)
```

```
##     RMSE      MAE     MAPE 
## 15.66667 15.66667 46.07843
```

```r
ro$predictions
```

```
##            h=1
## [1,] 49.66667
```

```r
ro$h_accu
```

```
##           h=1
## RMSE 15.66667
## MAE  15.66667
## MAPE 46.07843
```

```r
#for n = 3, k = 5, 7
pred <- knn_forecasting(knnt$Predictive.analytics...United.States., h=1, lags
=1:3, k =c(5,7))
pred$prediction
```

```
## Time Series:
## Start = 260 
## End = 260 
## Frequency = 1 
## [1] 43.14286
```

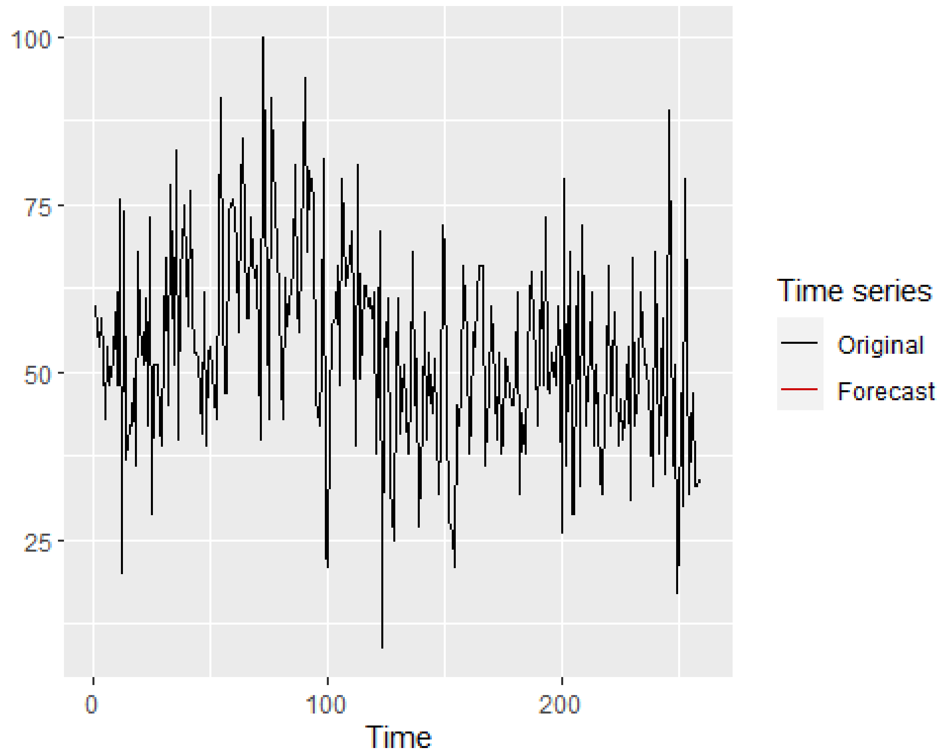
```r
pred$neighbors
```

```
## [1] 219 185  17 259 257 227 155
```

```r
nearest_neighbors(pred)
```

```
## $instance
## Lag 3 Lag 2 Lag 1 
##    33    33    34 
## 
## $nneighbors
##   Lag 3 Lag 2 Lag 1 H1
## 1    35    32    45 48
## 2    32    44    38 51
## 3    37    42    42 49
## 4    47    33    33 34
## 5    32    41    47 33
## 6    39    46    40 43
## 7    28    26    21 45
```

```r
#Plotting time series
autoplot(pred, h=1)
```




```r
#Calculating accuracy
ro <- rolling_origin(pred, h=1)
ro$global_accu #(Evaluating Using RMSE, MAE, MAPE)

##     RMSE      MAE     MAPE 
## 18.08571 18.08571 53.19328

ro$predictions

##           h=1
## [1,] 52.08571

ro$h_accu

##           h=1
## RMSE 18.08571
## MAE  18.08571
## MAPE 53.19328
```

Conclusion:

As we can see from the models results, the higher the dimension the model is, the lower the error we get.

References:

https://cran.r-project.org/web/packages/tsfknn/vignettes/tsfknn.html
https://cran.r-project.org/web/packages/philentropy/vignettes/Distances.html