

KNN MODEL

Santiago Orlando Moreno Suárez

2023-10-05

Practice on Predictive Analytics of Health Indicators and Diabetes

Introduction

In this practice, a predictive analysis on health indicators and diabetes is performed using the k-Nearest Neighbors (KNN) algorithm. The objective is to understand the relationship between various health indicators and the presence of diabetes. A detailed data exploration is carried out, the data is prepared for machine learning and a KNN model is trained to make predictions. The code is explained part by part below.

Part 1: Exploration and Data Manipulation

```
library(tidyverse)
library(caret)
library(MASS)
library(glmnet)
library(boot)
```

- `install.packages(c("caret", "MASS", "glmnet", "boot"))` installs the specified R packages if they are not already installed. These packages are essential for various data analysis tasks.
- `library(tidyverse)` loads the **tidyverse** package, which is a collection of R packages designed for data manipulation and visualization. It includes packages like **ggplot2** and **dplyr**.
- `library(caret)`, `library(MASS)`, `library(glmnet)`, and `library(boot)` load the respective packages into the R environment. These packages are used for machine learning, statistical modeling, and bootstrapping techniques.

Reading the Dataset:

```
data <- read.csv("diabetes_012_health_indicators_BRFSS2015.csv")
```

`read.csv("diabetes_012_health_indicators_BRFSS2015.csv")` reads the CSV (Comma-Separated Values) file named “diabetes_012_health_indicators_BRFSS2015.csv” into an R dataframe named **data**. This dataset likely contains health-related indicators, possibly including information related to diabetes, which will be used for further analysis in the subsequent parts of the code

Part 2: Data Sampling and Preparation

1. Setting Seed and Sampling Data:

```
set.seed(123)
sampled_data <- data %>% sample_frac(0.01)
```

- **set.seed(123)**: Sets the random seed to 123. This ensures that if you run the code again, you will get the same random results. It's useful for reproducibility in data analysis.
- **data %>% sample_frac(0.01)**: Takes a random 1% sample of the original dataset (**data**). The **%>%** operator, also known as the pipe operator, is used here to pipe the output of the previous command (**data**) into the **sample_frac** function, which then samples 1% of the data.

2. Creating Binary Diabetes Labels:

```
sampled_data$DiabetesBinary <- make.names(factor(ifelse(sampled_data$Diabetes_012 > 0, 1, 0)))
sampled_data$DiabetesBinary <- factor(sampled_data$DiabetesBinary, levels = c("X0", "X1"))
```

- **make.names(factor(ifelse(sampled_data\$Diabetes_012 > 0, 1, 0)))**: Converts the numeric variable **Diabetes_012** into a binary factor variable. If **Diabetes_012** is greater than 0, it's set to 1, otherwise 0. **factor** converts these numeric values into factors.
- **sampled_data\$DiabetesBinary <- factor(...)**: Assigns the resulting factors to a new column called **DiabetesBinary** in the **sampled_data** dataframe.
- **factor(..., levels = c("X0", "X1"))**: Specifies the levels of the factor. Here, "X0" represents the absence of diabetes (when **Diabetes_012** is 0) and "X1" represents the presence of diabetes (when **Diabetes_012** is greater than 0).

3. Creating Subsets:

```
sampled_data_bmi <- sampled_data
sampled_data_menthlth <- sampled_data
sampled_data_physhtlth <- sampled_data
```

These lines create three new dataframes (**sampled_data_bmi**, **sampled_data_menthlth**, and **sampled_data_physhtlth**), each containing the same data as **sampled_data**. These subsets are likely created to perform specific analyses or modeling tasks on different health indicators while keeping the original sampled data intact for reference.

Part 3: K-Nearest Neighbors (KNN) Classification

1. Setting Seed and Creating Training and Test Sets:

```
set.seed(123)
train_index <- createDataPartition(sampled_data$DiabetesBinary, p = 0.8, list = FALSE, times = 1)
train_data <- sampled_data[train_index, ]
test_data <- sampled_data[-train_index, ]
```

- `set.seed(123)`: Sets the random seed to 123 for reproducibility.
- `createDataPartition(sampled_data$DiabetesBinary, p = 0.8, list = FALSE, times = 1)`: Splits the `sampled_data` into training and test sets. 80% of the data is used for training (`train_data`), and 20% is used for testing (`test_data`). The indices for the training set are stored in `train_index`.

2. Setting up Control Parameters for KNN Model:

```
sampled_data_bmi <- sampled_data
sampled_data_menthlth <- sampled_data
sampled_data_physhtlth <- sampled_data
ctrl <- trainControl(method = "cv", number = 10, classProbs = TRUE, summaryFunction = twoClassSummary)
```

`trainControl(...)`: Configures the control parameters for the model training process. In this case, 10-fold cross-validation (`method = "cv"`) is used with class probabilities enabled (`classProbs = TRUE`). The `twoClassSummary` function is specified for summarizing the results for binary classification.

3. Training and Evaluating K-Nearest Neighbors (KNN) Model:

```
set.seed(20)
knn_model <- train(DiabetesBinary ~ ., data = train_data, method = "knn", trControl = ctrl, tuneLength = 10)
```

```
## Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not
## in the result set. ROC will be used instead.
```

```
predictions_knn <- predict(knn_model, newdata = test_data)
confusionMatrix(predictions_knn, test_data$DiabetesBinary)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  X0  X1
##           X0 425  51
##           X1   3  27
##
##              Accuracy : 0.8933
##              95% CI : (0.8631, 0.9188)
##    No Information Rate : 0.8458
##    P-Value [Acc > NIR] : 0.00128
##
##              Kappa : 0.4532
##
##  Mcnemar's Test P-Value : 1.596e-10
##
##              Sensitivity : 0.9930
##              Specificity : 0.3462
##    Pos Pred Value : 0.8929
##    Neg Pred Value : 0.9000
##    Prevalence : 0.8458
```

```
##          Detection Rate : 0.8399
##    Detection Prevalence : 0.9407
##          Balanced Accuracy : 0.6696
##
##          'Positive' Class : X0
##
```

- `set.seed(20)`: Sets a different random seed (20) for the KNN model training process.
- `train(...)`: Trains the KNN model. `DiabetesBinary ~ .` indicates that the column `DiabetesBinary` is the target variable, and all other columns are used as features. The training data is `train_data`, and the control parameters are specified by `trControl`.
- `predict(...)`: Uses the trained `knn_model` to make predictions on the test data (`test_data`).
- `confusionMatrix(...)`: Computes the confusion matrix to evaluate the performance of the KNN model by comparing the predicted values (`predictions_knn`) with the actual values (`test_data$DiabetesBinary`).

```
print(knn_model)
```

```
## k-Nearest Neighbors
##
## 2031 samples
## 22 predictor
## 2 classes: 'X0', 'X1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1827, 1828, 1827, 1828, 1827, 1829, ...
## Resampling results across tuning parameters:
##
##  k   ROC       Sens       Spec
##  5  0.9039012  0.9802020  0.4092742
##  7  0.9157860  0.9854413  0.3492944
##  9  0.9157347  0.9871855  0.3146169
## 11  0.9193260  0.9912621  0.2608871
## 13  0.9163338  0.9935979  0.2262097
## 15  0.9110241  0.9935945  0.1881048
## 17  0.9150622  0.9970862  0.1719758
## 19  0.9159526  0.9959200  0.1559476
## 21  0.9106427  0.9959200  0.1335685
## 23  0.9108658  0.9976676  0.1114919
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was k = 11.
```

```
knn_model
```

```
## k-Nearest Neighbors
##
## 2031 samples
## 22 predictor
```

```
## 2 classes: 'X0', 'X1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1827, 1828, 1827, 1828, 1827, 1829, ...
## Resampling results across tuning parameters:
##
## k ROC Sens Spec
## 5 0.9039012 0.9802020 0.4092742
## 7 0.9157860 0.9854413 0.3492944
## 9 0.9157347 0.9871855 0.3146169
## 11 0.9193260 0.9912621 0.2608871
## 13 0.9163338 0.9935979 0.2262097
## 15 0.9110241 0.9935945 0.1881048
## 17 0.9150622 0.9970862 0.1719758
## 19 0.9159526 0.9959200 0.1559476
## 21 0.9106427 0.9959200 0.1335685
## 23 0.9108658 0.9976676 0.1114919
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was k = 11.
```

`print(knn_model)`: This line of code prints the details of the `knn_model`, which was trained using the `train` function. When you print a trained model in R, it shows various information about the model, including its parameters, training performance, and tuning results (if applicable). This information is crucial for understanding the characteristics of the trained model.

Part 4: Linear and Multilinear Regression

1. Linear Regression Model for BMI:

```
model_bmi <- lm(BMI ~ ., data = sampled_data_bmi)
```

- `lm(BMI ~ ., data = sampled_data_bmi)`: Fits a linear regression model where `BMI` is the dependent variable, and `.` represents that all other columns in the `sampled_data_bmi` dataframe are used as independent variables. This means the model is trying to predict `BMI` based on other available variables in the dataset.

2. Cross-Validation for the BMI Regression Model:

```
cv_results_bmi <- cv.glm(data = sampled_data_bmi, glmfit = model_bmi, K = 10)
```

`cv.glm(data = sampled_data_bmi, glmfit = model_bmi, K = 10)`: Performs 10-fold cross-validation (`K = 10`) using the linear regression model (`glmfit = model_bmi`) on the data in `sampled_data_bmi`. Cross-validation is a technique used to assess how well a statistical model generalizes to an independent dataset. It involves dividing the dataset into `K` subsets, training the model on `K-1` of the folds, and testing it on the remaining fold. This process is repeated `K` times, with each of the folds used exactly once as the validation data.

3. Printing Cross-Validation Results:

```
print(cv_results_bmi)
```

```
## $call
## cv.glm(data = sampled_data_bmi, glmfit = model_bmi, K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 6 288629443 841063122 2108107620 1319689895
## [7] 1124255741 -1916009996 -1369100582 -372076507 126868799 2139108038
## [13] 1799511888 -1857306765 1380663697 660968384 -1769081442 1508296553
## [19] -1732164709 -426792278 1191347404 375635343 1324292549 1724341628
## [25] 1174234578 1517904205 802890759 -146189330 1285618376 1266070987
## [31] -986189783 -1142548872 698135462 564008385 -475090349 1291958530
## [37] 68857908 -1225209417 2119179693 433354436 -577414422 721526101
## [43] -757050641 1987293430 -1512222688 -1032932893 2025640577 1835220208
## [49] 1720578958 1961643257 546076299 -462125382 493488572 1565817599
## [55] -1518682539 -546780820 -573089342 259936797 -131544425 1812202494
## [61] 1939070904 -194817125 -1871185415 8522056 -478464906 -1395203663
## [67] -685682717 1460718770 -964919036 457599559 -1960932067 -109855852
## [73] -976097478 -1387654267 -1891137441 -1151354138 -1788296848 -1674360173
## [79] 143520561 105624032 1416218238 -1077891831 996507451 -206629110
## [85] -803272980 1426322287 -734427995 729757276 285359666 421600557
## [91] 1236110311 -289641714 1953434856 -313383189 -824686711 -1711246568
## [97] -713308602 -2037557599 -1361594061 198088866 1587222420 -49891817
## [103] -1992269299 -1266125020 1597977354 2035334773 863239631 -1935137066
## [109] -1665815936 1243026627 1075827297 -293492016 258329646 -380100135
## [115] 384351723 -1346447270 -896443876 1578253087 -809832587 -724552180
## [121] -1498976286 38868157 769725239 -1916503074 1948399896 -998044549
## [127] -924569063 -1858817368 1189501014 1500558545 542232707 -677125358
## [133] 1550808740 1827545703 2031860669 -311085772 -619606374 691215589
## [139] -150741377 -692811770 1668492176 -1860045261 708273617 1448785792
## [145] -1296192930 -1180003927 -1983774629 327669866 1130576140 1459147599
## [151] 31297541 -2125320516 532260754 -330115955 -1863150905 1400898478
## [157] 416783880 611193355 339493609 -688558664 1907952230 960706305
## [163] 1633665427 297491010 -118673932 160913399 -656258067 -1500289916
## [169] -436028758 54833173 1416231087 640521910 -1713904416 -1220270045
## [175] -307316927 -463422928 -1474451378 1885788345 -456061749 1032769914
## [181] -1112781700 1730914495 149675413 -613966420 -1205110910 1787813213
## [187] -1551707049 -867716290 582163320 219273307 -666640455 -862953976
## [193] 3722422 584211441 -362764765 745199218 906300612 328125063
## [199] 8602205 1472840788 -1489187206 359394757 1290096415 1350646438
## [205] 608173104 1423400659 582696177 1222924320 -977174850 -808166711
## [211] -1753534341 -1433391542 -767468884 -1682387537 -1067909787 -82973924
## [217] 629346162 -818109459 2063623207 -581567410 -302366552 2078731947
## [223] 1624837065 241412056 784608902 272316257 -1211667725 -1155091102
## [229] -1966849068 1677638130 -1166415568 1198840996 -1062057288 153580378
```

##	[235]	1807848848	-254656132	1777333524	1181233474	-591447296	1197300668
##	[241]	625075024	202791202	-571605688	-1050802164	-690356132	-222024494
##	[247]	1674496000	1044413268	-833459736	1193582586	-1949999408	786396460
##	[253]	-715960076	1413576594	1244240992	970353932	-15827232	-1872437918
##	[259]	1490924200	795339340	-866695300	-1718781134	-45472144	670879044
##	[265]	512842712	577511546	-370887120	-845491524	-591681708	-1845145022
##	[271]	-1234414688	-1079496516	1575242192	1682453858	-206783512	1990836172
##	[277]	-53045508	522936306	1644981376	554825972	-275508088	-2069109446
##	[283]	1473880912	-519959444	-139751276	-145639086	-1209334048	1041957068
##	[289]	1141649568	-1849937534	-934121240	-1255186196	1270071356	-934264718
##	[295]	-184768400	-1705223644	-824262088	-228928998	1407285456	-604432836
##	[301]	3675284	-1238457214	1594503104	-1860807364	504063632	307018786
##	[307]	-744470520	832124044	2012399644	1705477266	775142528	-1689404396
##	[313]	-205712600	-2087236550	-307086128	1730006124	-379029836	1955078162
##	[319]	2070937568	1693250508	1486271968	-1345930206	390788392	-641441524
##	[325]	1204051516	-706095886	-1734895376	1198786372	1561657432	-372606406
##	[331]	1200089328	722531004	-1735241196	787713218	50701920	-1433193092
##	[337]	-535409456	143760418	1438124200	2122333580	1683564988	336353330
##	[343]	1617394496	-39713100	-1419300536	-1307237702	752015120	-343242516
##	[349]	-1831724972	1414262546	202506656	829542348	588437344	-537070398
##	[355]	1341611560	722405804	-293448900	1279669106	655267632	1439852580
##	[361]	-258245064	1313002202	1915168656	-2007077636	-1746479724	-1842015422
##	[367]	758598144	1378527292	472083024	386957474	-362900664	-804068852
##	[373]	368653020	-463804334	-1513345920	1156247380	-798261272	1308490490
##	[379]	507068240	1368560684	1219326324	-1863846254	-2130306464	1882413196
##	[385]	-559085216	2042194146	791536808	1657036492	-1390876164	936817586
##	[391]	-1337000976	-1256576572	-290771624	-1877206662	267065776	-1263017540
##	[397]	-1849946540	1525791170	-745616608	-137045060	744616016	-1728797726
##	[403]	1572456296	-723165364	-2021355780	-1027993870	-780088064	-999398284
##	[409]	-89559544	1772584378	750945744	-479678484	-331838572	-1292750638
##	[415]	1120517216	210859340	-884503776	-1938556286	1412702696	1830773740
##	[421]	-1479403972	-1887453582	-1827540112	1225641252	-415376840	1179095450
##	[427]	-253442608	-1196366404	-1050851692	290816642	-953477696	1043345212
##	[433]	-223317360	1900799778	-1624222200	-1632664820	1319754652	-1382284270
##	[439]	783274880	-754382060	108508456	-1042771910	79770576	1226889836
##	[445]	1880505396	264837266	2034123104	-1878872756	-1868476960	3978146
##	[451]	-1378481240	777014412	-387677252	-1079788430	202070000	-1583737788
##	[457]	-428316803	-41876025	-1389813584	-38658354	-1574149125	2001593437
##	[463]	433159050	1887751080	-726906607	-49872893	-925072636	201113394
##	[469]	-223666809	-478903343	849845590	-783133900	738430533	1959336591
##	[475]	-825551864	-1641725914	-126500653	-51482059	-60272238	-1576594080
##	[481]	-1764623159	1163427835	-1434189876	-698281382	169209679	775626265
##	[487]	-1405019090	-1691057604	1370292141	-1263749033	-13640288	447516414
##	[493]	-451408437	-221013043	303091450	-625334984	776802017	-82812525
##	[499]	1804814772	1618238146	-1926649129	-1197384927	-1057932570	-1757889564
##	[505]	-180372715	1344447295	-1479457064	-2055355658	1308654851	1014519365
##	[511]	657649250	-1735662128	913265401	-1236557781	-1141149796	1567162314
##	[517]	1259927743	-446502391	781034270	909560524	23003805	-1692053721
##	[523]	1254113616	-357666834	1068986523	1012693757	1465229802	1871088264
##	[529]	1760435249	855830179	1042876068	2056800082	-284234073	-328199439
##	[535]	-1579025866	1947670228	-1769686427	1735579759	1646918312	-2053453818
##	[541]	631645491	-963909227	-1669581326	-652146112	-827989079	72482331
##	[547]	-245884948	11943930	1978000431	2055492665	226035278	1240102556
##	[553]	1464323085	-611757961	-970925952	1070718686	96089515	833488557

```
## [559] -1659643110 -1004747816 -1700454335 992759539 829042964 -161168222
## [565] -1030872649 1473570177 -1041927290 1048723908 -2089015051 1491921247
## [571] 350495672 -1455815018 1934069923 942414437 2137804674 -1452718864
## [577] -131933031 973599883 -1974796932 1616661418 105352159 1320935785
## [583] -1653879426 -1052160660 -1580214851 1410324103 -422301072 -352511730
## [589] -1455552965 581883165 587448266 -1615099672 113413073 2140990915
## [595] -643881276 1207512434 -1533968313 -734985199 1216275862 1740896500
## [601] 1110291077 -409633841 -542461240 -1512139930 -120919661 331735797
## [607] -1594230062 1087333664 1275710729 -1192060357 -1164865652 -801763814
## [613] -1179519345 1494946265 -1842960786 -740358660 -137494675 -1534875369
## [619] 77235424 -1842713922 84404363 1761744653 547501882 -1263857800
## [625] -1671551711 1109158558
```

```
cv_results_bmi
```

```
## $call
## cv.glm(data = sampled_data_bmi, glmfit = model_bmi, K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 6 288629443 841063122 2108107620 1319689895
## [7] 1124255741 -1916009996 -1369100582 -372076507 126868799 2139108038
## [13] 1799511888 -1857306765 1380663697 660968384 -1769081442 1508296553
## [19] -1732164709 -426792278 1191347404 375635343 1324292549 1724341628
## [25] 1174234578 1517904205 802890759 -146189330 1285618376 1266070987
## [31] -986189783 -1142548872 698135462 564008385 -475090349 1291958530
## [37] 68857908 -1225209417 2119179693 433354436 -577414422 721526101
## [43] -757050641 1987293430 -1512222688 -1032932893 2025640577 1835220208
## [49] 1720578958 1961643257 546076299 -462125382 493488572 1565817599
## [55] -1518682539 -546780820 -573089342 259936797 -131544425 1812202494
## [61] 1939070904 -194817125 -1871185415 8522056 -478464906 -1395203663
## [67] -685682717 1460718770 -964919036 457599559 -1960932067 -109855852
## [73] -976097478 -1387654267 -1891137441 -1151354138 -1788296848 -1674360173
## [79] 143520561 105624032 1416218238 -1077891831 996507451 -206629110
## [85] -803272980 1426322287 -734427995 729757276 285359666 421600557
## [91] 1236110311 -289641714 1953434856 -313383189 -824686711 -1711246568
## [97] -713308602 -2037557599 -1361594061 198088866 1587222420 -49891817
## [103] -1992269299 -1266125020 1597977354 2035334773 863239631 -1935137066
## [109] -1665815936 1243026627 1075827297 -293492016 258329646 -380100135
## [115] 384351723 -1346447270 -896443876 1578253087 -809832587 -724552180
## [121] -1498976286 38868157 769725239 -1916503074 1948399896 -998044549
## [127] -924569063 -1858817368 1189501014 1500558545 542232707 -677125358
## [133] 1550808740 1827545703 2031860669 -311085772 -619606374 691215589
## [139] -150741377 -692811770 1668492176 -1860045261 708273617 1448785792
## [145] -1296192930 -1180003927 -1983774629 327669866 1130576140 1459147599
## [151] 31297541 -2125320516 532260754 -330115955 -1863150905 1400898478
## [157] 416783880 611193355 339493609 -688558664 1907952230 960706305
## [163] 1633665427 297491010 -118673932 160913399 -656258067 -1500289916
## [169] -436028758 54833173 1416231087 640521910 -1713904416 -1220270045
```


## [175]	-307316927	-463422928	-1474451378	1885788345	-456061749	1032769914
## [181]	-1112781700	1730914495	149675413	-613966420	-1205110910	1787813213
## [187]	-1551707049	-867716290	582163320	219273307	-666640455	-862953976
## [193]	3722422	584211441	-362764765	745199218	906300612	328125063
## [199]	8602205	1472840788	-1489187206	359394757	1290096415	1350646438
## [205]	608173104	1423400659	582696177	1222924320	-977174850	-808166711
## [211]	-1753534341	-1433391542	-767468884	-1682387537	-1067909787	-82973924
## [217]	629346162	-818109459	2063623207	-581567410	-302366552	2078731947
## [223]	1624837065	241412056	784608902	272316257	-1211667725	-1155091102
## [229]	-1966849068	1677638130	-1166415568	1198840996	-1062057288	153580378
## [235]	1807848848	-254656132	1777333524	1181233474	-591447296	1197300668
## [241]	625075024	202791202	-571605688	-1050802164	-690356132	-222024494
## [247]	1674496000	1044413268	-833459736	1193582586	-1949999408	786396460
## [253]	-715960076	1413576594	1244240992	970353932	-15827232	-1872437918
## [259]	1490924200	795339340	-866695300	-1718781134	-45472144	670879044
## [265]	512842712	577511546	-370887120	-845491524	-591681708	-1845145022
## [271]	-1234414688	-1079496516	1575242192	1682453858	-206783512	1990836172
## [277]	-53045508	522936306	1644981376	554825972	-275508088	-2069109446
## [283]	1473880912	-519959444	-139751276	-145639086	-1209334048	1041957068
## [289]	1141649568	-1849937534	-934121240	-1255186196	1270071356	-934264718
## [295]	-184768400	-1705223644	-824262088	-228928998	1407285456	-604432836
## [301]	3675284	-1238457214	1594503104	-1860807364	504063632	307018786
## [307]	-744470520	832124044	2012399644	1705477266	775142528	-1689404396
## [313]	-205712600	-2087236550	-307086128	1730006124	-379029836	1955078162
## [319]	2070937568	1693250508	1486271968	-1345930206	390788392	-641441524
## [325]	1204051516	-706095886	-1734895376	1198786372	1561657432	-372606406
## [331]	1200089328	722531004	-1735241196	787713218	50701920	-1433193092
## [337]	-535409456	143760418	1438124200	2122333580	1683564988	336353330
## [343]	1617394496	-39713100	-1419300536	-1307237702	752015120	-343242516
## [349]	-1831724972	1414262546	202506656	829542348	588437344	-537070398
## [355]	1341611560	722405804	-293448900	1279669106	655267632	1439852580
## [361]	-258245064	1313002202	1915168656	-2007077636	-1746479724	-1842015422
## [367]	758598144	1378527292	472083024	386957474	-362900664	-804068852
## [373]	368653020	-463804334	-1513345920	1156247380	-798261272	1308490490
## [379]	507068240	1368560684	1219326324	-1863846254	-2130306464	1882413196
## [385]	-559085216	2042194146	791536808	1657036492	-1390876164	936817586
## [391]	-1337000976	-1256576572	-290771624	-1877206662	267065776	-1263017540
## [397]	-1849946540	1525791170	-745616608	-137045060	744616016	-1728797726
## [403]	1572456296	-723165364	-2021355780	-1027993870	-780088064	-999398284
## [409]	-89559544	1772584378	750945744	-479678484	-331838572	-1292750638
## [415]	1120517216	210859340	-884503776	-1938556286	1412702696	1830773740
## [421]	-1479403972	-1887453582	-1827540112	1225641252	-415376840	1179095450
## [427]	-253442608	-1196366404	-1050851692	290816642	-953477696	1043345212
## [433]	-223317360	1900799778	-1624222200	-1632664820	1319754652	-1382284270
## [439]	783274880	-754382060	108508456	-1042771910	79770576	1226889836
## [445]	1880505396	264837266	2034123104	-1878872756	-1868476960	3978146
## [451]	-1378481240	777014412	-387677252	-1079788430	202070000	-1583737788
## [457]	-428316803	-41876025	-1389813584	-38658354	-1574149125	2001593437
## [463]	433159050	1887751080	-726906607	-49872893	-925072636	201113394
## [469]	-223666809	-478903343	849845590	-783133900	738430533	1959336591
## [475]	-825551864	-1641725914	-126500653	-51482059	-60272238	-1576594080
## [481]	-1764623159	1163427835	-1434189876	-698281382	169209679	775626265
## [487]	-1405019090	-1691057604	1370292141	-1263749033	-13640288	447516414
## [493]	-451408437	-221013043	303091450	-625334984	776802017	-82812525

```
## [499] 1804814772 1618238146 -1926649129 -1197384927 -1057932570 -1757889564
## [505] -180372715 1344447295 -1479457064 -2055355658 1308654851 1014519365
## [511] 657649250 -1735662128 913265401 -1236557781 -1141149796 1567162314
## [517] 1259927743 -446502391 781034270 909560524 23003805 -1692053721
## [523] 1254113616 -357666834 1068986523 1012693757 1465229802 1871088264
## [529] 1760435249 855830179 1042876068 2056800082 -284234073 -328199439
## [535] -1579025866 1947670228 -1769686427 1735579759 1646918312 -2053453818
## [541] 631645491 -963909227 -1669581326 -652146112 -827989079 72482331
## [547] -245884948 11943930 1978000431 2055492665 226035278 1240102556
## [553] 1464323085 -611757961 -970925952 1070718686 96089515 833488557
## [559] -1659643110 -1004747816 -1700454335 992759539 829042964 -161168222
## [565] -1030872649 1473570177 -1041927290 1048723908 -2089015051 1491921247
## [571] 350495672 -1455815018 1934069923 942414437 2137804674 -1452718864
## [577] -131933031 973599883 -1974796932 1616661418 105352159 1320935785
## [583] -1653879426 -1052160660 -1580214851 1410324103 -422301072 -352511730
## [589] -1455552965 581883165 587448266 -1615099672 113413073 2140990915
## [595] -643881276 1207512434 -1533968313 -734985199 1216275862 1740896500
## [601] 1110291077 -409633841 -542461240 -1512139930 -120919661 331735797
## [607] -1594230062 1087333664 1275710729 -1192060357 -1164865652 -801763814
## [613] -1179519345 1494946265 -1842960786 -740358660 -137494675 -1534875369
## [619] 77235424 -1842713922 84404363 1761744653 547501882 -1263857800
## [625] -1671551711 1109158558
```

`print(cv_results_bmi)`: Prints the results of cross-validation. This likely includes metrics such as mean squared error, mean absolute error, or other relevant statistics that indicate how well the regression model performs on the validation sets during the cross-validation process. Printing these results helps in understanding the performance of the model and comparing it against other models or variations of the same model.

1. Linear Regression Model for Mental Health (MentHlth):

```
model_menthlth <- lm(MentHlth ~ ., data = sampled_data_menthlth)
```

`lm(MentHlth ~ ., data = sampled_data_menthlth)`: This line of code creates a linear regression model where `MentHlth` is the dependent variable, and `.` indicates that all other columns in the `sampled_data_menthlth` dataframe are used as independent variables. The model aims to predict `MentHlth` based on other available variables in the dataset.

2. Cross-Validation for the Mental Health Regression Model:

```
cv_results_menthlth <- cv.glm(data = sampled_data_menthlth, glmfit = model_menthlth, K = 10)
```

`cv.glm(data = sampled_data_menthlth, glmfit = model_menthlth, K = 10)`: This line performs 10-fold cross-validation ($K = 10$) using the linear regression model (`glmfit = model_menthlth`) on the data in the `sampled_data_menthlth` dataframe. Cross-validation is a technique used to assess how well a statistical model generalizes to an independent dataset. It involves splitting the dataset into K subsets, training the model on $K-1$ of the folds, and testing it on the remaining fold. This process is repeated K times, with each fold used exactly once as the validation data.

3. Printing Cross-Validation Results:

```
print(cv_results_menthlth)
```

```
## $call
## cv.glm(data = sampled_data_menthlth, glmfit = model_menthlth,
##       K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 546 -1264927468 2080692621 -1402187640 1051932561
## [7] 1360857648 689231512 1395983871 -213944491 -2017144420 212223559
## [13] 868708814 2121988964 -162915693 917324465 60358971 675043596
## [19] -871288111 1877386376 -1145208328 1756795925 -1574152591 -24902659
## [25] 814931698 -789501326 -1117129360 -628548098 -994879439 -1738675579
## [31] -988489512 224991277 2085062118 1350327312 378131532 961371695
## [37] -67081728 1548469910 773769810 -200416204 865672874 563619840
## [43] -386079820 -2073535396 1826289087 1203451461 1646486066 126806767
## [49] 6146615 567968577 1078292418 2023713473 1819366991 1416670708
## [55] -416389554 -1619392465 -1233331685 745284731 -2115296489 -110137260
## [61] 1417775928 796981555 -155321447 -1407989165 -2005283247 -2015986876
## [67] 645831705 1507887950 749479958 -1189870406 -691954523 -407069124
## [73] -608847280 230452693 -346657948 -160199428 1681693206 1197677007
## [79] 1381538823 -2051019093 -1864346757 -109810074 -1443438571 -292837820
## [85] 817266531 -653905635 -773293584 -431817941 -303024293 264335087
## [91] -2121560075 1542408963 1119360980 -844047285 4851489 -920161915
## [97] 1412839143 -1175977334 -1117550386 1821721943 1601548211 684285404
## [103] -388484940 561978347 452685008 -37202889 -1087033495 661553826
## [109] 911325662 -553188337 -2133980409 969824934 -1196542184 -1179115456
## [115] -126569757 1840771492 585725691 -1298190275 1809623248 -74500826
## [121] 284349037 1670130493 -1498717528 1336884079 -1488213859 1129197098
## [127] 460712394 -1361344432 254868894 -797234809 1100115318 2096063949
## [133] -1615037485 1162129113 326980552 -1857078259 -195922473 920287172
## [139] 1828071945 1150318639 -581825635 1502670086 -795973958 -1552503242
## [145] 1164439064 -802375939 -532354068 -1253137733 1763349800 -1995995946
## [151] -1707145765 131471402 -1903270454 2036498450 495094374 -1634032248
## [157] 1293391091 -1206980492 847436292 -621189176 -1485158238 1987083883
## [163] 510943313 1586877512 -582589730 -656502251 -1847150743 -1760676790
## [169] 1384836717 43404332 -1986202972 987791550 -1499964212 -1854297988
## [175] -1634669541 472485267 -1008764749 1285170536 -1932971039 -419378491
## [181] 818000038 1173771287 -379656551 -151490502 339038621 45575052
## [187] 413375703 683818333 -565056469 -1062894953 848798046 -1886122473
## [193] -655091611 374695017 225655680 176485802 493925246 1149420733
## [199] 1817054961 -1770910751 -79328517 -1028992841 -2092504688 -789749549
## [205] -161451857 -1275130683 -835306290 -1848645387 -945400475 -296048843
## [211] -891671614 -1752772778 529722765 1108570416 1213285330 -1040816736
## [217] 611497142 -1616978473 -1784192157 -1401326618 464989343 -1136570845
## [223] 1584794490 1653377665 -352892425 -833539857 -1909336497 -1251461873
```

##	[229]	257090475	-58637746	-909236999	1505146242	-207121482	1900541609
##	[235]	-2103359683	-211155720	30453545	-777921778	2058942049	2130260338
##	[241]	1242842121	-1556331367	289955738	-1195520083	-1789726583	-1110338103
##	[247]	159491427	948574333	218668341	-1484278627	-971760526	1631052515
##	[253]	1583368456	-2084072307	-1598237902	-203050390	-1837487731	1719835696
##	[259]	-933783703	373545642	-1647511561	1519830719	-480462587	899117120
##	[265]	-1243528599	2098558062	1554893213	-1015682455	-1021852150	-544377949
##	[271]	-1481899148	878895622	491613701	351943648	411383205	-2101367202
##	[277]	-455679782	383071406	1089753928	-845564051	652349506	-60104882
##	[283]	1220428304	-270751832	-2126419766	-1292606469	-1969976292	902822691
##	[289]	-1526256060	-827571176	457601660	-1985921828	679728549	1438688917
##	[295]	859622344	-1294262502	-233051510	-728948870	877965150	-1576775114
##	[301]	232373295	618450303	11582906	-850947610	524969763	2052648236
##	[307]	2031483923	-1877519229	-1267037149	1427074962	-430370226	1865001113
##	[313]	-1164549825	1196015814	152370423	-205557815	-576875724	-2089309764
##	[319]	1821633067	1326762222	1896031773	591096925	2116500182	426465752
##	[325]	1035675871	-281746216	560903270	-548628982	72016401	-1464797355
##	[331]	-75492104	1231689299	-1192965168	445750334	-1932852685	-2116863764
##	[337]	-1989238023	-50520279	-721422940	-750255632	-701639028	410650061
##	[343]	2143103289	1412770959	-999849912	-758066235	995590855	1876485298
##	[349]	-564768692	-1496236144	1727319319	-1714485028	-687193873	26752631
##	[355]	216114853	1264264197	19766430	2074642850	2114131355	1153550249
##	[361]	1506361597	461250667	46559561	-1276697983	479678852	-816416038
##	[367]	-1913466585	1490534381	1471730454	-865853613	1582630270	2129815140
##	[373]	-911236960	942025889	-171710211	1939298558	-1564169526	690630867
##	[379]	-2011563706	-595539966	-1425390863	1563793613	-139170984	929791589
##	[385]	921452258	2032468542	956654713	-2394581	-1608397942	495990443
##	[391]	1826419595	2076461603	-1212871650	-1785143492	2067869989	217841864
##	[397]	1055166603	917724703	793681326	-1600149959	-1593843702	329409930
##	[403]	-1558860947	-486126913	-1996297833	621071822	87910122	-62202202
##	[409]	1211179257	-551088016	1144731956	795590817	1076758443	516702359
##	[415]	-7073966	483273689	854399066	-92851538	918376434	-496204726
##	[421]	699062451	794395251	-230868826	1083553664	711019120	-1056514632
##	[427]	935054644	-1631900884	989776121	541661625	-1067429784	-270293036
##	[433]	-81430835	579268902	2104854681	-2033903971	1022499717	87856472
##	[439]	1920129732	-1582550821	1985843773	452676056	736078257	20197195
##	[445]	-1728040254	-32328370	661524273	-32608348	743312959	-1993127109
##	[451]	-1604408620	-1865248119	-1961682705	-1114062264	-814381652	-1380695799
##	[457]	1850681278	2060892760	1432766947	-856986159	181187425	562825521
##	[463]	-952411033	1752889766	1784186966	-1846926964	-1858201950	-1111647044
##	[469]	718327358	1299132996	-1470339627	1320318743	1941295721	206587813
##	[475]	123790496	-227010721	771050806	335413405	2061489969	916191224
##	[481]	952040272	-2119894426	701748328	61096146	1835930738	-2104747050
##	[487]	-1235286455	-965678212	1865270093	-1086168539	-1675843541	912824435
##	[493]	-963254110	1964264316	582842935	988547141	-1837780430	-1176095522
##	[499]	2022828830	315287472	934188946	-917205049	81185014	-54626070
##	[505]	1623625179	-2075001663	1742879337	1572059245	-818831125	1731495905
##	[511]	275023318	622875575	-2014808016	1392001499	647971070	-1200839448
##	[517]	236320735	-911318208	-360686413	-1341204674	-1638568786	-338907020
##	[523]	-1682176480	-1986322402	-2144854423	-626212731	-2030234316	1313768003
##	[529]	-1190580648	-1207378734	1274072327	-1194959759	-201410254	976874914
##	[535]	-1576599336	735478536	-740991007	-2136314511	1487568543	489917976
##	[541]	-26636740	-1171252129	-1035190912	-1308676535	795040118	1329858902
##	[547]	-1409222401	-285985179	2007837820	221596702	-2126247272	823085981

```
## [553] -1289129857 1308032511 1527856052 -1547794572 542944064 -2082073494
## [559] -892629344 1056759262 1234847468 -871719068 -1994037637 -1960479955
## [565] -1132044614 373165975 320066521 1302316372 -1049886017 -855379743
## [571] 1140437682 -2062326876 -640024301 -2061992490 1145031800 1448778692
## [577] -186148326 78566755 -665153354 908542671 1549609867 -209584585
## [583] 604047555 1293839596 366347173 -1126558241 -1717657089 1050066599
## [589] 1920860082 186628473 1891117450 1265376276 2005488615 -328068709
## [595] -276296322 1620306223 -128671129 879288182 907085425 -1312718875
## [601] 1292608701 -1293166349 -1490042787 5126014 -2123383556 -386217624
## [607] 2006802521 -803985363 -448887100 1519239978 -1696669237 2141758422
## [613] -344574172 915262315 752326672 1566103007 -961225924 1015176344
## [619] 1312893845 164758663 -750666113 -1355804395 1094476023 -481271833
## [625] -347392048 1963180836
```

```
cv_results_menthlth
```

```
## $call
## cv.glm(data = sampled_data_menthlth, glmfit = model_menthlth,
##       K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 546 -1264927468 2080692621 -1402187640 1051932561
## [7] 1360857648 689231512 1395983871 -213944491 -2017144420 212223559
## [13] 868708814 2121988964 -162915693 917324465 60358971 675043596
## [19] -871288111 1877386376 -1145208328 1756795925 -1574152591 -24902659
## [25] 814931698 -789501326 -1117129360 -628548098 -994879439 -1738675579
## [31] -988489512 224991277 2085062118 1350327312 378131532 961371695
## [37] -67081728 1548469910 773769810 -200416204 865672874 563619840
## [43] -386079820 -2073535396 1826289087 1203451461 1646486066 126806767
## [49] 6146615 567968577 1078292418 2023713473 1819366991 1416670708
## [55] -416389554 -1619392465 -1233331685 745284731 -2115296489 -110137260
## [61] 1417775928 796981555 -155321447 -1407989165 -2005283247 -2015986876
## [67] 645831705 1507887950 749479958 -1189870406 -691954523 -407069124
## [73] -608847280 230452693 -346657948 -160199428 1681693206 1197677007
## [79] 1381538823 -2051019093 -1864346757 -109810074 -1443438571 -292837820
## [85] 817266531 -653905635 -773293584 -431817941 -303024293 264335087
## [91] -2121560075 1542408963 1119360980 -844047285 4851489 -920161915
## [97] 1412839143 -1175977334 -1117550386 1821721943 1601548211 684285404
## [103] -388484940 561978347 452685008 -37202889 -1087033495 661553826
## [109] 911325662 -553188337 -2133980409 969824934 -1196542184 -1179115456
## [115] -126569757 1840771492 585725691 -1298190275 1809623248 -74500826
## [121] 284349037 1670130493 -1498717528 1336884079 -1488213859 1129197098
## [127] 460712394 -1361344432 254868894 -797234809 1100115318 2096063949
## [133] -1615037485 1162129113 326980552 -1857078259 -195922473 920287172
## [139] 1828071945 1150318639 -581825635 1502670086 -795973958 -1552503242
## [145] 1164439064 -802375939 -532354068 -1253137733 1763349800 -1995995946
## [151] -1707145765 131471402 -1903270454 2036498450 495094374 -1634032248
## [157] 1293391091 -1206980492 847436292 -621189176 -1485158238 1987083883
```

## [163]	510943313	1586877512	-582589730	-656502251	-1847150743	-1760676790
## [169]	1384836717	43404332	-1986202972	987791550	-1499964212	-1854297988
## [175]	-1634669541	472485267	-1008764749	1285170536	-1932971039	-419378491
## [181]	818000038	1173771287	-379656551	-151490502	339038621	45575052
## [187]	413375703	683818333	-565056469	-1062894953	848798046	-1886122473
## [193]	-655091611	374695017	225655680	176485802	493925246	1149420733
## [199]	1817054961	-1770910751	-79328517	-1028992841	-2092504688	-789749549
## [205]	-161451857	-1275130683	-835306290	-1848645387	-945400475	-296048843
## [211]	-891671614	-1752772778	529722765	1108570416	1213285330	-1040816736
## [217]	611497142	-1616978473	-1784192157	-1401326618	464989343	-1136570845
## [223]	1584794490	1653377665	-352892425	-833539857	-1909336497	-1251461873
## [229]	257090475	-58637746	-909236999	1505146242	-207121482	1900541609
## [235]	-2103359683	-211155720	30453545	-777921778	2058942049	2130260338
## [241]	1242842121	-1556331367	289955738	-1195520083	-1789726583	-1110338103
## [247]	159491427	948574333	218668341	-1484278627	-971760526	1631052515
## [253]	1583368456	-2084072307	-1598237902	-203050390	-1837487731	1719835696
## [259]	-933783703	373545642	-1647511561	1519830719	-480462587	899117120
## [265]	-1243528599	2098558062	1554893213	-1015682455	-1021852150	-544377949
## [271]	-1481899148	878895622	491613701	351943648	411383205	-2101367202
## [277]	-455679782	383071406	1089753928	-845564051	652349506	-60104882
## [283]	1220428304	-270751832	-2126419766	-1292606469	-1969976292	902822691
## [289]	-1526256060	-827571176	457601660	-1985921828	679728549	1438688917
## [295]	859622344	-1294262502	-233051510	-728948870	877965150	-1576775114
## [301]	232373295	618450303	11582906	-850947610	524969763	2052648236
## [307]	2031483923	-1877519229	-1267037149	1427074962	-430370226	1865001113
## [313]	-1164549825	1196015814	152370423	-205557815	-576875724	-2089309764
## [319]	1821633067	1326762222	1896031773	591096925	2116500182	426465752
## [325]	1035675871	-281746216	560903270	-548628982	72016401	-1464797355
## [331]	-75492104	1231689299	-1192965168	445750334	-1932852685	-2116863764
## [337]	-1989238023	-50520279	-721422940	-750255632	-701639028	410650061
## [343]	2143103289	1412770959	-999849912	-758066235	995590855	1876485298
## [349]	-564768692	-1496236144	1727319319	-1714485028	-687193873	26752631
## [355]	216114853	1264264197	19766430	2074642850	2114131355	1153550249
## [361]	1506361597	461250667	46559561	-1276697983	479678852	-816416038
## [367]	-1913466585	1490534381	1471730454	-865853613	1582630270	2129815140
## [373]	-911236960	942025889	-171710211	1939298558	-1564169526	690630867
## [379]	-2011563706	-595539966	-1425390863	1563793613	-139170984	929791589
## [385]	921452258	2032468542	956654713	-2394581	-1608397942	495990443
## [391]	1826419595	2076461603	-1212871650	-1785143492	2067869989	217841864
## [397]	1055166603	917724703	793681326	-1600149959	-1593843702	329409930
## [403]	-1558860947	-486126913	-1996297833	621071822	87910122	-62202202
## [409]	1211179257	-551088016	1144731956	795590817	1076758443	516702359
## [415]	-7073966	483273689	854399066	-92851538	918376434	-496204726
## [421]	699062451	794395251	-230868826	1083553664	711019120	-1056514632
## [427]	935054644	-1631900884	989776121	541661625	-1067429784	-270293036
## [433]	-81430835	579268902	2104854681	-2033903971	1022499717	87856472
## [439]	1920129732	-1582550821	1985843773	452676056	736078257	20197195
## [445]	-1728040254	-32328370	661524273	-32608348	743312959	-1993127109
## [451]	-1604408620	-1865248119	-1961682705	-1114062264	-814381652	-1380695799
## [457]	1850681278	2060892760	1432766947	-856986159	181187425	562825521
## [463]	-952411033	1752889766	1784186966	-1846926964	-1858201950	-1111647044
## [469]	718327358	1299132996	-1470339627	1320318743	1941295721	206587813
## [475]	123790496	-227010721	771050806	335413405	2061489969	916191224
## [481]	952040272	-2119894426	701748328	61096146	1835930738	-2104747050

```
## [487] -1235286455 -965678212 1865270093 -1086168539 -1675843541 912824435
## [493] -963254110 1964264316 582842935 988547141 -1837780430 -1176095522
## [499] 2022828830 315287472 934188946 -917205049 81185014 -54626070
## [505] 1623625179 -2075001663 1742879337 1572059245 -818831125 1731495905
## [511] 275023318 622875575 -2014808016 1392001499 647971070 -1200839448
## [517] 236320735 -911318208 -360686413 -1341204674 -1638568786 -338907020
## [523] -1682176480 -1986322402 -2144854423 -626212731 -2030234316 1313768003
## [529] -1190580648 -1207378734 1274072327 -1194959759 -201410254 976874914
## [535] -1576599336 735478536 -740991007 -2136314511 1487568543 489917976
## [541] -26636740 -1171252129 -1035190912 -1308676535 795040118 1329858902
## [547] -1409222401 -285985179 2007837820 221596702 -2126247272 823085981
## [553] -1289129857 1308032511 1527856052 -1547794572 542944064 -2082073494
## [559] -892629344 1056759262 1234847468 -871719068 -1994037637 -1960479955
## [565] -1132044614 373165975 320066521 1302316372 -1049886017 -855379743
## [571] 1140437682 -2062326876 -640024301 -2061992490 1145031800 1448778692
## [577] -186148326 78566755 -665153354 908542671 1549609867 -209584585
## [583] 604047555 1293839596 366347173 -1126558241 -1717657089 1050066599
## [589] 1920860082 186628473 1891117450 1265376276 2005488615 -328068709
## [595] -276296322 1620306223 -128671129 879288182 907085425 -1312718875
## [601] 1292608701 -1293166349 -1490042787 5126014 -2123383556 -386217624
## [607] 2006802521 -803985363 -448887100 1519239978 -1696669237 2141758422
## [613] -344574172 915262315 752326672 1566103007 -961225924 1015176344
## [619] 1312893845 164758663 -750666113 -1355804395 1094476023 -481271833
## [625] -347392048 1963180836
```

`print(cv_results_menthlth)`: This line prints the results of the cross-validation. The output likely includes metrics such as mean squared error, mean absolute error, or other relevant statistics indicating how well the regression model performs on the validation sets during the cross-validation process. Printing these results helps in understanding the performance of the model and comparing it against other models or variations of the same model.

1. Linear Regression Model for Physical Health (PhysHlth):

```
model_physhtlth <- lm(PhysHlth ~ ., data = sampled_data_physhtlth)
```

`lm(PhysHlth ~ ., data = sampled_data_physhtlth)`: This line creates a linear regression model where `PhysHlth` is the dependent variable, and `.` indicates that all other columns in the `sampled_data_physhtlth` dataframe are used as independent variables. The model attempts to predict `PhysHlth` based on other available variables in the dataset.

2. Cross-Validation for the Physical Health Regression Model:

```
cv_results_physhtlth <- cv.glm(data = sampled_data_physhtlth, glmfit = model_physhtlth, K = 10)
```

`cv.glm(data = sampled_data_physhtlth, glmfit = model_physhtlth, K = 10)`: This line performs 10-fold cross-validation ($K = 10$) using the linear regression model (`glmfit = model_physhtlth`) on the data in the `sampled_data_physhtlth` dataframe. Cross-validation is a technique used to assess how well a statistical model generalizes to an independent dataset. It involves splitting the dataset into K subsets, training the model on $K-1$ of the folds, and testing it on the remaining fold. This process is repeated K times, with each fold used exactly once as the validation data.

```
print(cv_results_physhtlth)
```

```
## $call
## cv.glm(data = sampled_data_physhtlth, glmfit = model_physhtlth,
##       K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 501 -998478309 2025147153 2131157516 1204875709
## [7] 1959715502 1182780599 796517853 -1817230012 1285445053 -1899384932
## [13] -1314656260 1272204161 1889447085 -1784616947 1737451913 770648557
## [19] 1236183864 1997280833 527607085 1126068770 -1628920983 -1951159852
## [25] 2064061845 1656602352 794462428 1242380941 -73466082 -115043679
## [31] 939738958 -1599620813 -720237235 1785615374 1288265728 486592706
## [37] 1796713881 -622632796 1828427312 -1469495344 -1041347894 2089277803
## [43] 813392690 -1011843565 1288560573 -1957008884 -1187519603 398474302
## [49] -169954062 1613231702 -38790703 -2142332733 -1637099173 1210872453
## [55] 1841142182 109063323 -448589162 2116770180 -1795621760 -1441692541
## [61] 522410678 -579489970 -1200842413 -1228555658 1095354718 -1290679244
## [67] -1624166260 242224723 -1975919011 1335496747 -344597652 1263821819
## [73] 1815993903 -322287352 -1746015457 -510923338 1660802637 19770543
## [79] -80910556 256157290 -347389221 -954026246 240541538 -857961287
## [85] -1803960513 3703704 630893462 -1911468568 570982575 1832369319
## [91] -462143038 1247124959 -1451382790 1532181102 280291345 1822242783
## [97] -598380845 1891915384 -23328871 89249742 1221812228 1775208746
## [103] 1775228664 -1647401725 1017779391 -252689468 147857450 -1295585257
## [109] 1325134541 978047133 1429707310 -294832307 2011246126 655566960
## [115] -321808914 2000431545 -2045897255 904223037 102550336 340494428
## [121] 940622909 -1172383097 2053468830 1601795976 -808300489 -529643345
## [127] 2082430313 -560977124 -1240632600 31358980 -1571064858 -1706717396
## [133] -65609398 1320004490 2007167890 1249340956 649681396 -273105127
## [139] 149150953 2035566605 173286835 565755996 -986894433 -1809416630
## [145] 1100916103 1719830102 -46523287 113902241 -1167107057 -1866626254
## [151] -1002780612 -935870504 -212562488 -1726885523 1948831935 -367367588
## [157] -916013425 -1006705633 -1124659019 1745492326 428936654 1262982641
## [163] -929923713 222810966 -1535268753 536752583 -411199313 1322245858
## [169] -132158684 -305864412 141247791 -30968821 1738142648 459947390
## [175] 2086054022 1469306051 -918606634 373255002 857966834 1766337141
## [181] 423764047 974964491 111068683 837601620 801694517 -1855672990
## [187] -1367727875 1427163945 -414142704 -37686764 1334989933 1580939884
## [193] -1069648183 357991895 1395537880 -1925388494 -907453089 915332154
## [199] 1328072092 -1743823139 160142437 1001645698 1591855616 -637239910
## [205] -965714717 1049372254 1630764713 421666538 2136705345 1399123596
## [211] 182079049 1026477111 474003794 -1589187435 139616714 893671196
## [217] -1110338906 -1317816429 -616474131 -1271959258 1573953715 289615931
## [223] 1753979814 591609410 -1818831321 1589006197 260886123 165524416
## [229] 177565777 1703734685 -702058554 -1192605509 2019773205 -1713552024
## [235] -933539794 586031024 -1480338402 442280300 1901531300 718894502
```


##	[241]	1766281313	505048272	-607674105	593136559	-407802404	-1686122461
##	[247]	558965166	-1701006689	406068393	-1284415217	-1624245677	-1060326016
##	[253]	-2020619472	564678577	1301781193	1202639008	1737363001	-1103815540
##	[259]	1948075807	646285343	698934286	-1292882837	-2003203937	-44042735
##	[265]	-1058919857	1755948037	1137388390	341593695	1872884052	-110981322
##	[271]	1580402468	-1937574083	1012259508	114176170	-1240586333	-2033670696
##	[277]	-2090920211	266208324	-909440145	-1624359777	-61545148	-449158809
##	[283]	642819752	-1952726297	-1369184825	-2097190673	-615582050	1192346432
##	[289]	-1907785692	1589045756	78882008	-757211631	-978647018	281060233
##	[295]	-1374895942	-2501230	1750731029	-1266200078	221618012	-907829101
##	[301]	-1256251628	-1613587418	-1019323809	892893	417084335	1351401389
##	[307]	1158577416	-567751664	537201205	622511263	1221003058	-1374107624
##	[313]	1842669554	1964580503	1226814617	-215348534	1067119157	607062695
##	[319]	-144604575	-685020628	1127813048	692421424	-1392676226	1379714780
##	[325]	-1061679278	1588811326	-773763516	1585500910	1920509310	556978075
##	[331]	198884590	783573980	1079791981	701268441	-977045654	1342024105
##	[337]	-1288105039	-400088518	863522232	301617132	1083846368	658768336
##	[343]	-944931094	1318575037	-1559505230	-188521759	-63913222	-645435047
##	[349]	-1174286429	1209329922	-1441875784	604023813	483941821	2075477082
##	[355]	609109386	1253422111	-327595522	1352026654	-1571292980	-337254933
##	[361]	1305477384	26372338	-1051664876	-288788200	-20482293	-25358076
##	[367]	942106019	845927404	-1332104664	-1780043465	1013746935	1421026155
##	[373]	-1515502348	1471208261	-2137137748	-1018152234	-1739683387	-743872501
##	[379]	961319260	-1711959121	-290536663	2083790526	382798857	-938509017
##	[385]	170003767	-151962305	-1516248015	631894160	-478149486	1145705623
##	[391]	-865896439	745428947	2043345426	-834801253	-418949747	-1443667181
##	[397]	1129178792	1554437208	1779005032	1277142370	-894785666	-2061065821
##	[403]	-1360437107	-235045944	1824487543	1832504259	-208234608	452995715
##	[409]	1714819645	287927173	967006695	1586352889	1819482186	-1040431843
##	[415]	675544578	741524375	-1478633308	421309761	1020572286	533493827
##	[421]	-1764574600	212814114	-91443612	-1381880280	-963979845	1038384637
##	[427]	932865957	836951892	1853029172	-288504739	-1039338923	199310244
##	[433]	544046500	1073380723	-1162038530	198293052	77471077	516423235
##	[439]	-1560765230	866160649	786460437	-1874163330	-1287550332	-295136612
##	[445]	1447838304	2128985443	-1950273609	1523428133	1415174992	-43904362
##	[451]	-620213375	-742040407	255649884	1861352143	1076396175	-1099919074
##	[457]	-181974384	435212003	-1524811706	439120801	15664871	679186064
##	[463]	716743696	-1529047668	122365841	-971220848	1565395977	-412273630
##	[469]	-294946237	-8249168	-1672694912	-1578594336	-71247605	-955886051
##	[475]	-1701036685	1112086024	1571584998	852589197	1652500125	-1213383157
##	[481]	143002464	-1784178818	-705924076	-934750663	-2116077565	-330356082
##	[487]	66807588	885546520	-1622007044	-203316835	1574556198	-1189305264
##	[493]	1734521121	1289023542	984122411	1735616168	2008039961	1379090908
##	[499]	591667894	1682435671	331575079	213228201	-537792561	1791961127
##	[505]	-1065749047	883618444	886274057	1838083148	-1357882070	-1405846877
##	[511]	517958902	1537961265	-1484152154	207259358	-1313776744	-657908797
##	[517]	84360050	-850048570	-1307869054	-266535351	-1715836965	-277067835
##	[523]	-1707715421	1582329853	-1229313130	-1383975408	532608038	-144269409
##	[529]	-1018439400	1800386104	565681379	427719473	2116265355	-1410492757
##	[535]	1113756103	292226428	1677463615	-740880857	459908723	164019395
##	[541]	-631762738	-862410766	1556786980	-110610764	1371882960	-1654965308
##	[547]	-1762428969	-1865920762	185993071	-1950058028	-132928798	620209382
##	[553]	-2102234119	1904195692	-138419897	-1917501514	1537893205	-327842054
##	[559]	1122219498	-940990357	881409177	-1114118330	-1786270451	934508717

```
## [565] -1271901040 -930098015 2093321118 -1741514922 1222095501 -1178420316
## [571] 592046389 -1199085784 1597218344 -476045283 901241950 1455479181
## [577] 1756600986 -1987809208 -1203696448 -664073389 204856569 422455163
## [583] -176088583 82630881 -744022096 -625540429 152655939 -420964319
## [589] -1443381192 -1017548672 1779474616 -504978206 1846653632 -715677775
## [595] 913394132 -1202138378 -782140976 -890801668 1383549765 -1622005559
## [601] -1579151093 412536910 -542547928 -1302793589 1368237795 -1025228941
## [607] 475567846 113412215 664840757 -1534897440 875250163 150322589
## [613] 1072884084 905330064 -1579619868 438421973 -699096526 1238865398
## [619] -52935918 99205894 -1170179280 -733501692 784964080 1475719067
## [625] 1518038160 -1858080326
```

```
cv_results_physhtlth
```

```
## $call
## cv.glm(data = sampled_data_physhtlth, glmfit = model_physhtlth,
##       K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] NaN NaN
##
## $seed
## [1] 10403 501 -998478309 2025147153 2131157516 1204875709
## [7] 1959715502 1182780599 796517853 -1817230012 1285445053 -1899384932
## [13] -1314656260 1272204161 1889447085 -1784616947 1737451913 770648557
## [19] 1236183864 1997280833 527607085 1126068770 -1628920983 -1951159852
## [25] 2064061845 1656602352 794462428 1242380941 -73466082 -115043679
## [31] 939738958 -1599620813 -720237235 1785615374 1288265728 486592706
## [37] 1796713881 -622632796 1828427312 -1469495344 -1041347894 2089277803
## [43] 813392690 -1011843565 1288560573 -1957008884 -1187519603 398474302
## [49] -169954062 1613231702 -38790703 -2142332733 -1637099173 1210872453
## [55] 1841142182 109063323 -448589162 2116770180 -1795621760 -1441692541
## [61] 522410678 -579489970 -1200842413 -1228555658 1095354718 -1290679244
## [67] -1624166260 242224723 -1975919011 1335496747 -344597652 1263821819
## [73] 1815993903 -322287352 -1746015457 -510923338 1660802637 19770543
## [79] -80910556 256157290 -347389221 -954026246 240541538 -857961287
## [85] -1803960513 3703704 630893462 -1911468568 570982575 1832369319
## [91] -462143038 1247124959 -1451382790 1532181102 280291345 1822242783
## [97] -598380845 1891915384 -23328871 89249742 1221812228 1775208746
## [103] 1775228664 -1647401725 1017779391 -252689468 147857450 -1295585257
## [109] 1325134541 978047133 1429707310 -294832307 2011246126 655566960
## [115] -321808914 2000431545 -2045897255 904223037 102550336 340494428
## [121] 940622909 -1172383097 2053468830 1601795976 -808300489 -529643345
## [127] 2082430313 -560977124 -1240632600 31358980 -1571064858 -1706717396
## [133] -65609398 1320004490 2007167890 1249340956 649681396 -273105127
## [139] 149150953 2035566605 173286835 565755996 -986894433 -1809416630
## [145] 1100916103 1719830102 -46523287 113902241 -1167107057 -1866626254
## [151] -1002780612 -935870504 -212562488 -1726885523 1948831935 -367367588
## [157] -916013425 -1006705633 -1124659019 1745492326 428936654 1262982641
## [163] -929923713 222810966 -1535268753 536752583 -411199313 1322245858
## [169] -132158684 -305864412 141247791 -30968821 1738142648 459947390
```

## [175]	2086054022	1469306051	-918606634	373255002	857966834	1766337141
## [181]	423764047	974964491	111068683	837601620	801694517	-1855672990
## [187]	-1367727875	1427163945	-414142704	-37686764	1334989933	1580939884
## [193]	-1069648183	357991895	1395537880	-1925388494	-907453089	915332154
## [199]	1328072092	-1743823139	160142437	1001645698	1591855616	-637239910
## [205]	-965714717	1049372254	1630764713	421666538	2136705345	1399123596
## [211]	182079049	1026477111	474003794	-1589187435	139616714	893671196
## [217]	-1110338906	-1317816429	-616474131	-1271959258	1573953715	289615931
## [223]	1753979814	591609410	-1818831321	1589006197	260886123	165524416
## [229]	177565777	1703734685	-702058554	-1192605509	2019773205	-1713552024
## [235]	-933539794	586031024	-1480338402	442280300	1901531300	718894502
## [241]	1766281313	505048272	-607674105	593136559	-407802404	-1686122461
## [247]	558965166	-1701006689	406068393	-1284415217	-1624245677	-1060326016
## [253]	-2020619472	564678577	1301781193	1202639008	1737363001	-1103815540
## [259]	1948075807	646285343	698934286	-1292882837	-2003203937	-44042735
## [265]	-1058919857	1755948037	1137388390	341593695	1872884052	-110981322
## [271]	1580402468	-1937574083	1012259508	114176170	-1240586333	-2033670696
## [277]	-2090920211	266208324	-909440145	-1624359777	-61545148	-449158809
## [283]	642819752	-1952726297	-1369184825	-2097190673	-615582050	1192346432
## [289]	-1907785692	1589045756	78882008	-757211631	-978647018	281060233
## [295]	-1374895942	-2501230	1750731029	-1266200078	221618012	-907829101
## [301]	-1256251628	-1613587418	-1019323809	892893	417084335	1351401389
## [307]	1158577416	-567751664	537201205	622511263	1221003058	-1374107624
## [313]	1842669554	1964580503	1226814617	-215348534	1067119157	607062695
## [319]	-144604575	-685020628	1127813048	692421424	-1392676226	1379714780
## [325]	-1061679278	1588811326	-773763516	1585500910	1920509310	556978075
## [331]	198884590	783573980	1079791981	701268441	-977045654	1342024105
## [337]	-1288105039	-400088518	863522232	301617132	1083846368	658768336
## [343]	-944931094	1318575037	-1559505230	-188521759	-63913222	-645435047
## [349]	-1174286429	1209329922	-1441875784	604023813	483941821	2075477082
## [355]	609109386	1253422111	-327595522	1352026654	-1571292980	-337254933
## [361]	1305477384	26372338	-1051664876	-288788200	-20482293	-25358076
## [367]	942106019	845927404	-1332104664	-1780043465	1013746935	1421026155
## [373]	-1515502348	1471208261	-2137137748	-1018152234	-1739683387	-743872501
## [379]	961319260	-1711959121	-290536663	2083790526	382798857	-938509017
## [385]	170003767	-151962305	-1516248015	631894160	-478149486	1145705623
## [391]	-865896439	745428947	2043345426	-834801253	-418949747	-1443667181
## [397]	1129178792	1554437208	1779005032	1277142370	-894785666	-2061065821
## [403]	-1360437107	-235045944	1824487543	1832504259	-208234608	452995715
## [409]	1714819645	287927173	967006695	1586352889	1819482186	-1040431843
## [415]	675544578	741524375	-1478633308	421309761	1020572286	533493827
## [421]	-1764574600	212814114	-91443612	-1381880280	-963979845	1038384637
## [427]	932865957	836951892	1853029172	-288504739	-1039338923	199310244
## [433]	544046500	1073380723	-1162038530	198293052	77471077	516423235
## [439]	-1560765230	866160649	786460437	-1874163330	-1287550332	-295136612
## [445]	1447838304	2128985443	-1950273609	1523428133	1415174992	-43904362
## [451]	-620213375	-742040407	255649884	1861352143	1076396175	-1099919074
## [457]	-181974384	435212003	-1524811706	439120801	15664871	679186064
## [463]	716743696	-1529047668	122365841	-971220848	1565395977	-412273630
## [469]	-294946237	-8249168	-1672694912	-1578594336	-71247605	-955886051
## [475]	-1701036685	1112086024	1571584998	852589197	1652500125	-1213383157
## [481]	143002464	-1784178818	-705924076	-934750663	-2116077565	-330356082
## [487]	66807588	885546520	-1622007044	-203316835	1574556198	-1189305264
## [493]	1734521121	1289023542	984122411	1735616168	2008039961	1379090908

```

## [499] 591667894 1682435671 331575079 213228201 -537792561 1791961127
## [505] -1065749047 883618444 886274057 1838083148 -1357882070 -1405846877
## [511] 517958902 1537961265 -1484152154 207259358 -1313776744 -657908797
## [517] 84360050 -850048570 -1307869054 -266535351 -1715836965 -277067835
## [523] -1707715421 1582329853 -1229313130 -1383975408 532608038 -144269409
## [529] -1018439400 1800386104 565681379 427719473 2116265355 -1410492757
## [535] 1113756103 292226428 1677463615 -740880857 459908723 164019395
## [541] -631762738 -862410766 1556786980 -110610764 1371882960 -1654965308
## [547] -1762428969 -1865920762 185993071 -1950058028 -132928798 620209382
## [553] -2102234119 1904195692 -138419897 -1917501514 1537893205 -327842054
## [559] 1122219498 -940990357 881409177 -1114118330 -1786270451 934508717
## [565] -1271901040 -930098015 2093321118 -1741514922 1222095501 -1178420316
## [571] 592046389 -1199085784 1597218344 -476045283 901241950 1455479181
## [577] 1756600986 -1987809208 -1203696448 -664073389 204856569 422455163
## [583] -176088583 82630881 -744022096 -625540429 152655939 -420964319
## [589] -1443381192 -1017548672 1779474616 -504978206 1846653632 -715677775
## [595] 913394132 -1202138378 -782140976 -890801668 1383549765 -1622005559
## [601] -1579151093 412536910 -542547928 -1302793589 1368237795 -1025228941
## [607] 475567846 113412215 664840757 -1534897440 875250163 150322589
## [613] 1072884084 905330064 -1579619868 438421973 -699096526 1238865398
## [619] -52935918 99205894 -1170179280 -733501692 784964080 1475719067
## [625] 1518038160 -1858080326

```

`print(cv_results_physhtlth)`: This line prints the results of the cross-validation. The output likely includes metrics such as mean squared error, mean absolute error, or other relevant statistics indicating how well the regression model performs on the validation sets during the cross-validation process. Printing these results helps in understanding the performance of the model and comparing it against other models or variations of the same model.

CONCLUSIONS:

1. **Data Preparation and Exploration:** The code starts by loading necessary libraries and reading the dataset. A 1% random sample of the data is taken for analysis. The dataset is then processed to create binary labels for diabetes (**DiabetesBinary**) and subsets of data are created for different health indicators (**BMI**, **MentHlth**, **PhysHlth**).
2. **Classification Using K-Nearest Neighbors (KNN):** A KNN classification model is trained to predict diabetes (**DiabetesBinary**). The dataset is split into training and test sets (80% - 20%). The KNN model is evaluated using 10-fold cross-validation. Various **k** values from 1 to 10 are tested, and the model's performance is assessed using confusion matrices.
3. **Regression Modeling for Health Indicators:** Separate linear regression models are built for predicting health indicators (**BMI**, **MentHlth**, **PhysHlth**). Cross-validation with 10 folds is performed for each regression model, and the mean squared error or other relevant metrics are used to evaluate model performance.
4. **Cross-Validation and Model Evaluation:** Cross-validation is extensively used to assess the models' performance, ensuring robustness and generalizability. The printed cross-validation results provide insights into the models' accuracy and help in choosing the best-performing models.
5. **Reproducibility and Seed Setting:** The code emphasizes reproducibility by setting random seeds (`set.seed(...)`) at critical points. Reproducibility is crucial in data analysis to ensure that results can be replicated, enhancing the credibility of the findings.

6. **Overall Analysis:** The code offers a comprehensive analysis of health indicators, employing both classification and regression techniques. By utilizing KNN for binary classification and linear regression for continuous prediction, the code provides a holistic understanding of the relationships between various health indicators and diabetes status.