Jayeeta Ghosh

From: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Sent: Thursday, August 1, 2019 1:53 PM
To: Jayeeta Ghosh
Cc: Jerry Hartanto
Subject: Re: Sample Data

Jayeeta,

The 32 features were given to me as is and normalized from patient records. That's all I got and had to make it work.

Sathiyan (Seth) Kutty
Principal - Predictive Analytics
Enterprise Change & Configuration Management
Kaiser Permanente
4460 Hacienda Dr, Pleasanton, CA 94588
email: Sathiyan.k.kutty@kp.org
cell: 650-888-8062

Upcoming PTO: 8/5-8/15

From: Jayeeta Ghosh <JGhosh@trace3.com>
Date: Thursday, August 1, 2019 at 1:32 PM
To: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Cc: Jerry Hartanto <jhartanto@trace3.com>
Subject: RE: Sample Data

Hello Seth,
Thank you very much for the dataset. I will surely play with it as soon as I get a chance. Any context you could provide on what exactly we are trying to predict here (bcva01 is the binary target variables) or data dictionary for 32 features that would be great. If not its okay, we will treat them as a challenge where we don't have lot of understanding of the features but rather just focus on the model.
Thank you very much.
Jayeeta

From: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Sent: Wednesday, July 31, 2019 3:18 PM
To: Jayeeta Ghosh <JGhosh@trace3.com>
Subject: Sample Data

Jayeeta,

Here's the dataset that you can test in your free time. If it works, I'll introduce you to Liyan Liu. She is a great partner to work with.

From: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Date: Friday, October 12, 2018 at 1:36 PM
To: Liyan Liu <liyan.liu@kp.org>
Subject: RE: Data Scientist Introductions

Liyan, all the models are hovering around 64-65%. So it looks like you have a winner! I will try next week; I am wrapping up a clustering model right now. I've attached the screenshots for you.

From: Liyan Liu
Sent: Friday, October 12, 2018 12:04 PM
To: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Subject: RE: Data Scientist Introductions

Now in csv format

First column BCVA01 is the outcome, all other columns are predictors

Thanks a ton,
Liyan

From: Sathiyan K Kutty
Sent: Friday, October 12, 2018 11:49 AM
To: Liyan Liu <Liyan.Liu@kp.org>
Subject: RE: Data Scientist Introductions

Liyan,

Can you send me the data in text format? I am not able to open the sas7bdat file.

Sent: Friday, October 12, 2018 10:49 AM
To: Sathiyan K Kutty <Sathiyan.K.Kutty@kp.org>
Subject: RE: Data Scientist Introductions

Dear Seth,

How are you doing?

At our last call, you mentioned / showed me your data science platform. I'm very impressed. I have used R package to run machine learning algorithms to predict visual acuity (variable name BCVA01). With gradient boosting, I got 70% AUC. Could you try these data on your platform to see if you can get better prediction? Thanks so much.

Attached are model building data rbuild and model testing data rtest.
Here are my R codes

```
##########################################
# Model 4: Gradient boosting: caret package
##########################################
```

```
# reload data
training <- read_sas("Rbuild.sas7bdat",NULL)
training$bcva01 <- ifelse(training$bcva01==1,'yes','nope')
training$bcva01 <- as.factor(training$bcva01)

testing <- read_sas("Rtest.sas7bdat",NULL)

# generalize variables so that I can recycle the codes for subsequent runs
outcomeName <- 'bcva01'
predictorsNames <- names(training)[names(training) != outcomeName]
predictorsNames

# model building
trCtrl <- trainControl(method='repeatedcv', number=5, returnResamp='none', summaryFunction = twoClassSummary,
classProbs = TRUE)

?train

Model_boost <- train(f, data=training,
                     method='gbm',
                     trControl=trCtrl,
                     metric = "ROC",
                     preProc = c("center", "scale"))

# Generate predictions on test dataset
p_gboost_caret <- predict(Model_boost, newdata = testing, type='prob')

p_gboost_caret

# Compute AUC on the test set
cvAUC::AUC(predictions = p_gboost_caret[,2], labels = testing[,"bcva01"]) #0.700
```

Thanks a ton and happy Friday!

With profound gratitude and respect,
Liyan

Liyan Liu, Msc.
Data Scientist
Division of Research, Kaiser permanente
2000 Broadway, Oakland, CA 94612
Liyan.liu@kp.org
5108913695

NOTICE TO RECIPIENT: If you are not the intended recipient of this e-mail, you are prohibited from sharing, copying, or otherwise using or disclosing its contents. If you have received this e-mail in error, please notify the sender immediately by reply e-mail and permanently delete this e-mail and any attachments without reading, forwarding or saving them. Thank you.