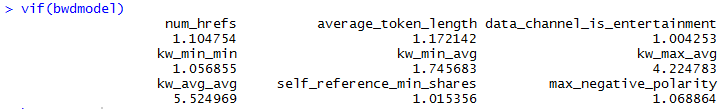
Andrés Felipe Paredes Jiménez

John Sebastián Marín Benítez

# Taller en clase: Multicolinealidad

Sí hay multicolinealidad.

Usando el método VIF obtenemos:



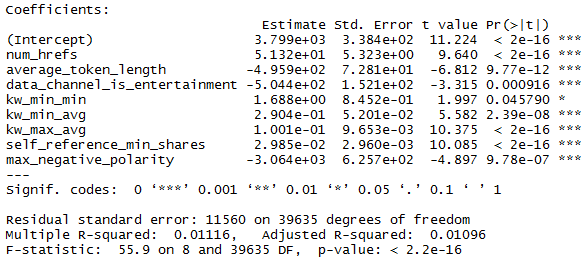
Podemos observar que hay un problema de multicolinealidad en las variables kw\_max\_avg y kw\_avg\_avg.

Usando el método Kappa obtenemos:



Este estadístico es muy grande por lo tanto hay un problema de multicolinealidad.

Ejecutando el método de eliminación del mayor VIF con criterio mayor a 4 obtenemos el siguiente modelo:



El método eliminó la variable kw\_avg\_avg.

Num\_hrefs, kw\_min\_min, kw\_min\_avg, kw\_max\_avg y self\_reference\_min\_shares afectan positivamente a la variable dependiente shares.

Average\_token\_lenght, data\_channel\_is\_entertainment y max\_negative\_polarity afectan negativamente a la variable dependiente shares.

El modelo explica el 1.12% de la variabilidad de shares.

library(car)

bwdmodel <- lm(shares ~ num\_hrefs + average\_token\_length + data\_channel\_is\_entertainment + kw\_min\_min +

kw\_min\_avg + kw\_max\_avg + kw\_avg\_avg + self\_reference\_min\_shares +

max\_negative\_polarity, data = data)

vif(bwdmodel)

#Esto nos dice que hay multicolinealidad en kw\_max\_avg, kw\_avg\_avg

XTX <- model.matrix(bwdmodel)

e <- eigen(t(XTX) %\*% XTX)

e$val

lambda.1 <- max(e$val)

lambda.k <- min(e$val)

kappa <- sqrt(lambda.1/lambda.k)

kappa

#219921.9 Este estadístico es muy grande por lo tanto hay multicolinealidad

remueve.VIF.grande <- function(modelo, u){

require(car)

# extrae el dataframe

data <- modelo$model

# Calcula todos los VIF

all\_vifs <- car::vif(modelo)

# extraer el nombre de todas las variables X

names\_all <- names(all\_vifs)

# extraer el nombre de la variables y

dep\_var <- all.vars(formula(modelo))[1]

# Remover lsa variables con VIF > u

# y reestimar el modelo con las otras variables

while(any(all\_vifs > u)){

# elimina variable con max vif

var\_max\_vif <- names(which(all\_vifs == max(all\_vifs)))

# remueve la variable

names\_all <- names\_all[!(names\_all) %in% var\_max\_vif]

# nueva formula

myForm <- as.formula(paste(paste(dep\_var, "~ "),

paste (names\_all, collapse=" + "), sep=""))

# re-build model with new formula

modelo.prueba <- lm(myForm, data= data)

all\_vifs <- car::vif(modelo.prueba)

}

modelo.limpio <- modelo.prueba

return(modelo.limpio)

}

modeloGenial<-remueve.VIF.grande(bwdmodel,4)

summary(modeloGenial)