

Compilation of all codes:

```
library(tidyr)

head(qsar_oral_toxicity)
names(qsar_oral_toxicity)
names(qsar_oral_toxicity) <- c("x1")

library(stringr)
####Here are all the codes used for the project

OralTox2 <- as.data.frame(str_split_fixed(qsar_oral_toxicity$x1, ";", 1025))

Unclassified <- OralTox2[-c(1025),drop=FALSE]
rowSums(sapply(No_Low_No_Intergenic_snpeff.scores[, c(20:29, 45)],
              function(x) as.numeric(as.character(x)))))

OralTox2["Sum"] <- rowSums(sapply(Unclassified[, c(1:1024)],
                                function(x) as.numeric(as.character(x)))))

#install.packages('e1701', dependencies=TRUE)
#install.packages('caret', dependencies=TRUE)
library(tidyr)

head(qsar_oral_toxicity)
names(qsar_oral_toxicity)
names(qsar_oral_toxicity) <- c("x1")

library(stringr)

OralTox <- as.data.frame(str_split_fixed(qsar_oral_toxicity$x1, ";", 1025))
OralTox$V1025 <- factor(OralTox$V1025, levels=c("negative","positive"))

split_row <- round(nrow(OralTox)*.60, 0)
train <- OralTox[1:split_row, ]
test <- OralTox[(split_row+1):nrow(OralTox), ]

model <- glm(V1025~., family="binomial", data=train)
summary(model)
p <- predict(model, newdata=test, type="response")
classes <- ifelse(p>=0.5, 1, 0)
p_class <- factor(classes, labels=c("negative","positive"))
```

```
confusionMatrix(data=p_class, reference=test$V1025)
```