**CENG 3522: Applied Machine Learning**

**PROJECT DOCUMENT**

**Project Name:**

NEWYORKCITY INSPECTIONS

**Team Members:**

ASLI YAVUZ

ÇAĞATAY ŞENTURCA

**Description of the Learning Problem:**

Classification problem in which we will identify the state of each provided restaurant, the elements which leaded to the grade and whether the future restaurants have the same possibilities.

**Goal /Motivation:**

We are aiming to find the possibility of success or fail of restaurants in NewYorkCity.

**DATA**

After we applied all the changes to our data which we got from Kaggle , we came up with 11 columns. We had to remove some columns which are related to each other. When all things done with the pre-processing we inserted the data into Weka and started to predict. Our data includes these columns at the last form of data.

Column names;

BORO

STREET

CUISINE DESCRIPTION

VIOLATION CODE

VIOLATION DESCRIPTION

CRITICAL FLAG

SCORE

GRADE

INSPECTION TYPE

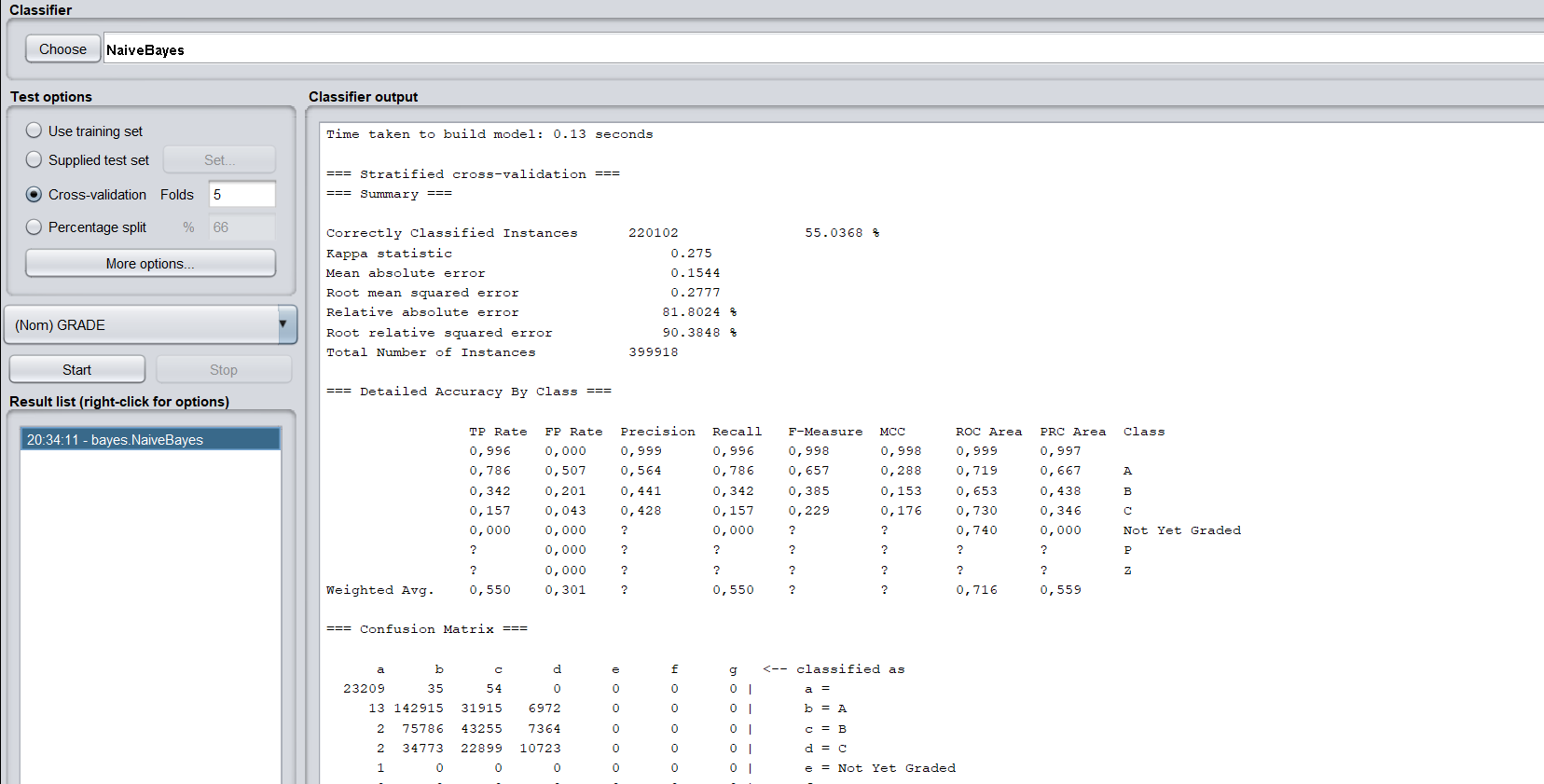
SEASON

RESTAURANT TYPE

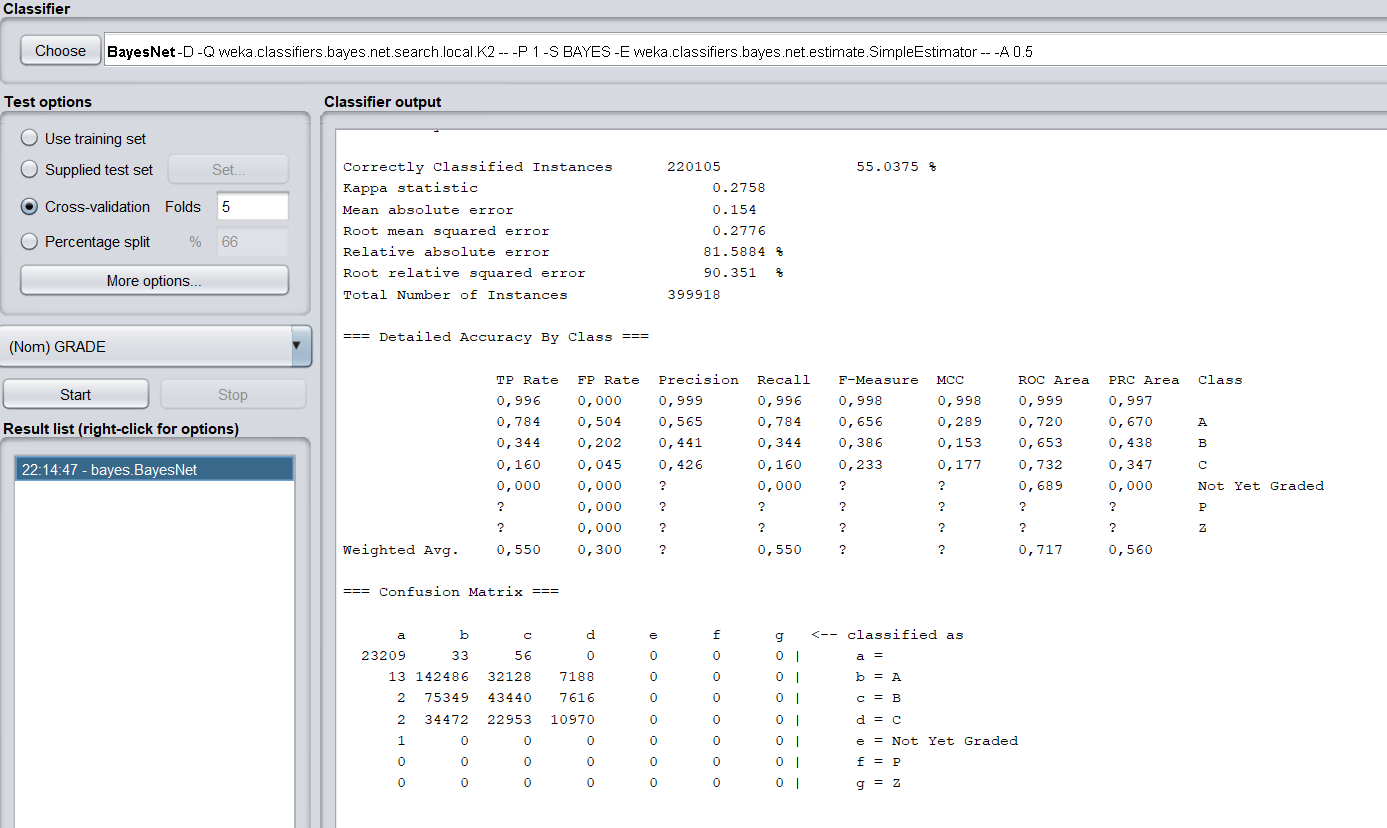
**RESULTS**

We predicted GRADE some algorithms.

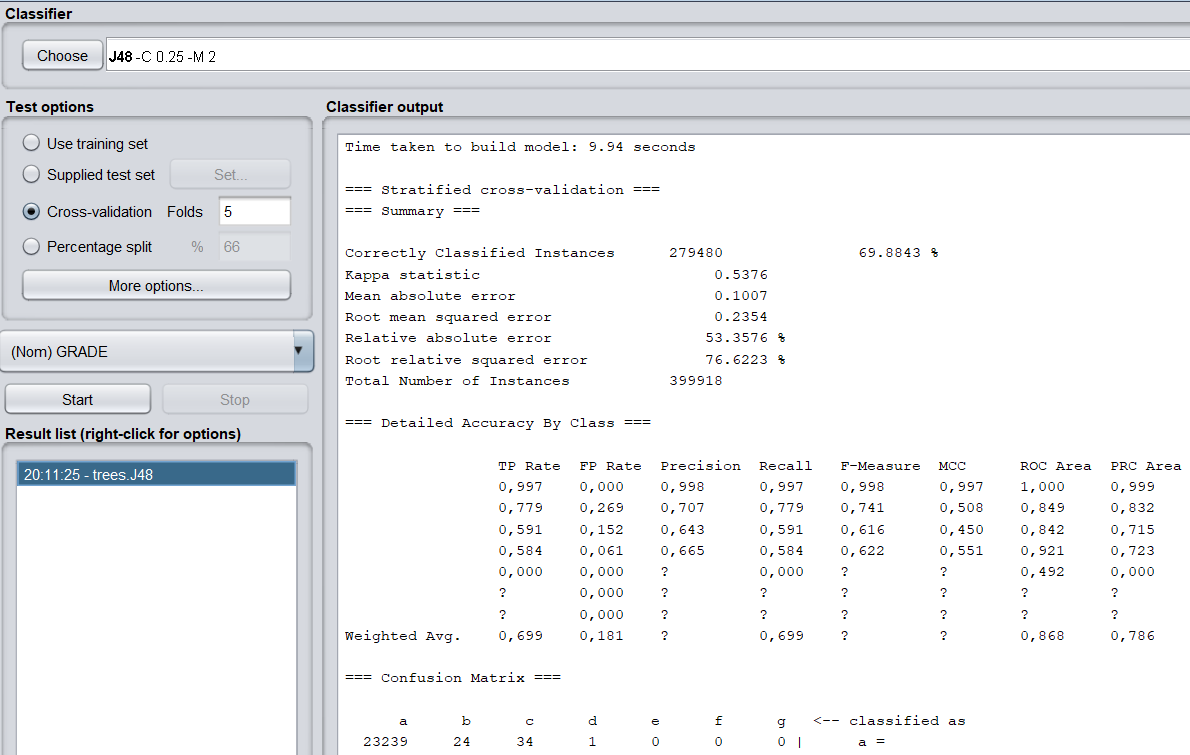
**NaiveBayes;**

****

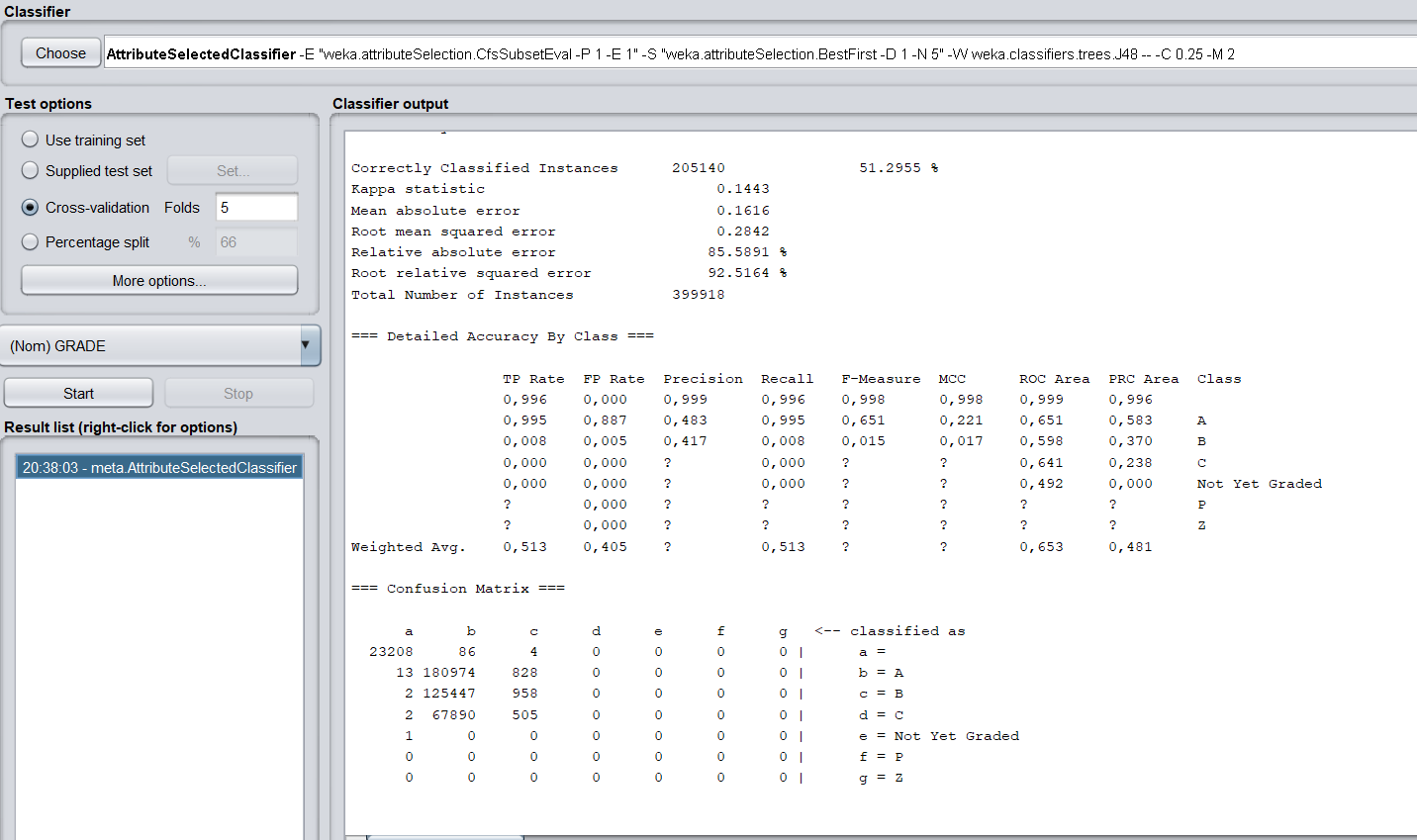
**BayesNet;**

****

**J48;**

****

**AttributeSelectedClassifier;**

****

After we applied algorithms on our dataset , as we expect that J48 algorithm gives the highest correctly classified instances score as percentage.

**SCRIPTS**

**R CODES;**

**FIRST FILE**

dir<-("/home/aslii/Desktop")

setwd(dir)

nyc <- read.csv("DOHMH\_New\_York\_City\_Restaurant\_Inspection\_Results.csv", header = TRUE)

nyc$INSPECTION.DATE <- as.Date(nyc$INSPECTION.DATE, format="%m/%d/%Y")

class(nyc$INSPECTION.DATE)

metseasons <- c(

"01" = "Winter", "02" = "Winter",

"03" = "Spring", "04" = "Spring", "05" = "Spring",

"06" = "Summer", "07" = "Summer", "08" = "Summer",

"09" = "Fall", "10" = "Fall", "11" = "Fall",

"12" = "Winter"

)

nyc$Season <- metseasons[format(nyc$INSPECTION.DATE, "%m")]

write.csv(nyc, file = "nyc.csv", row.names = FALSE)

**SECOND FILE**

dir<-("/home/aslii/Desktop")

setwd(dir)

library("foreign")

nyc1 <- read.csv("nyc.csv", header = TRUE)

nyc1$RESTAURANTTYPE<- ifelse(grepl("TAKE",nyc1$DBA,ignore.case = T),"TAKE OUT", "IN")

nyc1$RESTAURANTTYPE<-as.factor(nyc1$RESTAURANTTYPE)

nyc1$BUILDING <-as.numeric(nyc1$BUILDING)

nyc1$ZIPCODE <-as.numeric(nyc1$ZIPCODE)

nyc1$SCORE <-as.numeric(nyc1$SCORE)

na.omit(nyc1)

nyc1$GRADE[nyc1$SCORE >=0 & nyc1$SCORE <= 13 ]<- "A"

nyc1$GRADE[nyc1$SCORE >=14 & nyc1$SCORE <= 27]<- "B"

nyc1$GRADE[nyc1$SCORE >=28 ]<- "C"

which(as.numeric(nyc1$SCORE) == min(as.numeric(nyc1$SCORE),na.rm = TRUE))

nyc1 <- subset( nyc1, select = -c(1,2,4,6,7,9,10,16,17) )

write.arff(nyc1, file = "nycfinal.arff")