# Shiny App of Wine Classification using Predictive LDA Modeling

**Anuradha, Ayush Jhawar, Deepshikha, Rohit Verma**

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

[aanuradh@purdue.edu](mailto:aanuradh@purdue.edu), [ajhawar@purdue.edu](mailto:ajhawar@purdue.edu), [deepshik@purdue.edu](mailto:deepshik@purdue.edu), [verma68@purdue.edu](mailto:verma68@purdue.edu)

## Abstract

AADR wines aims to classify wines into categories based on their chemical content and identify customer segments that buy those categories of wines. Based on this information, future wines can be classified into these categories which can help in procurement activities as we will have a clearer idea of which wines our customers prefer. It will also help AADR to send targeted emails with better discount deals via email to customers. This will result in more emails being opened and fetch higher revenues.

The chemical content data of wines is used to classify the wines in 3 categories using LDA method. The predictive model is then tested on the test data. A shiny app is made with this predictive model running in the background. This app requires the user to upload the chemical data and displays the classification as output.

**Keywords:** Wine Classification, Targeted Advertising, LDA Modeling, Email Campaign, Predictive Modeling

## Business Problem

## Ayush and his partners own and run AADR Wines in West Lafayette. Ayush travels around the world, find the best deals on wines and ships them back to Lafayette. Back home, Anuradha, Deepshikha and Rohit have to sell these wines across the country. They reach out to customers via social media and emails. The email newsletters form the bulk of the business. The trio try to dig deeper into the sales data to figure out if there is a pattern in which the wines are bought. They decide to segment customers based on this data, so that they can send more targeted emails when the next batch comes in.

The stakeholders in this operation are the owners, the suppliers from whom the wine is purchased across the globe, and the customers who expect good quality product coming from AADR.

The different types of wines can be clustered based on their chemical content (alcohol, malic acid, magnesium, ash, hue, proline, etc.) and new wines can be allotted to these segments. These can then be promoted amongst people who prefer such kinds of wines.

Benefits of such a classification are that AADR can do targeted marketing and promote specific wines to specific groups of people via email which generates maximum business. This may result in more response to email campaigns and in turn higher revenue. An insight into such a classification will also help Ayush in purchasing the correct types of wines for customers and negotiate better deals.

This exercise is an internal process for the company and all partners agree on the business problem and the proposed solution.

## Analytics Problem

## The group tries to group wines together based on their characteristics and get 3 segments. They identify that particular types of customers who prefer one group of wines over the other. The problem is to make a model which accurately classifies a new wine into these 3 segments so that targeted emails can be sent to people in the geographies who prefer those wines.

The underlying assumption for this project is that the wines can be classified together based on their characteristics and that they can be matched to different customer segments.

Key metrics for success are predicting the type of wine correctly from the new data based on the model.

This exercise is an internal process for the company and all partners agree on the approach to the business problem.

## Data

The data from suppliers should necessarily contain the following information about the wine:

1. Alcohol
2. Malic acid
3. Ash
4. Alkalinity of ash
5. Magnesium
6. Total phenols
7. Flavanoids
8. Nonflavanoid phenols
9. Proanthocyanins
10. Color intensity
11. Hue
12. OD280/OD315 of diluted wines
13. Proline

Using the values for these parameters, new wines will be classified into the existing categories.

Missing values in any column would result in row deletion, which means that the wine will be disqualified from classification. The classification for that disqualified wine will have to be done manually.

The business and analytics problem do not need to be redefined.

## Methodology Selection

The team had two options to build the prediction model, Multinomial Logistic Regression and Linear Discriminant Analysis.

AADR is a fledgling company and cannot afford to use paid statistical software for such an activity. All the required features are available in R. Also, to make it easier for future classification, we can make an app which will do the classification and show us the results for an incoming batch of wines.

The team decided to go ahead with LDA.

## Model Building

The team got an accuracy of 98% using LDA to build the predictive model. The result was 3 classes of wines. On checking the buyers for these segments it was identified that 3 different geographies bought their wines differently. The states of Indiana and North Carolina preferred segment 1, Colorado and Texas preferred segment 2 and Ohio and Nebraska preferred segment 3. Preferences for other regions could not be established. The limitations to this exercise are that the model was built using 178 observations. This will be alleviated as more wines are classified and the data pool increases. Misclassification of certain wines is also a problem.

## Functionality

The Decision Support System (DSS) helps the user enter the details of the wine and classifies it based on the LDA predictive model. This helps the team send out relevant emails to customers containing deals to wines which they are more likely to buy.

Apart from the standard library functions, the **caret** and **shiny** packages were helpful in carrying out this project. The LDA modeling was performed with the help of the caret package. The shiny package helped in the creation of the app which makes it easier for the user to input the data and see the results.

After the LDA predictive model is made, the new data is classified based on the segments obtained from the initial data set.

Future enhancements for the project could include machine learning from the new data set being tested. This will improve the accuracy of the model.

## GUI Design and Functionality

The GUI design is easy to use and works intuitively. The user has to just upload the file with the relevant data in the form of a csv file and the app gives the classification results. It derives some look and feel from other Shiny apps we found on the web.

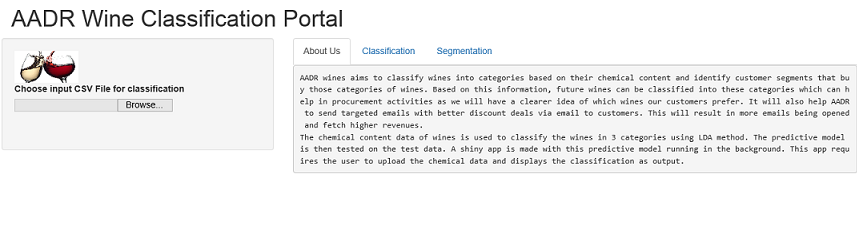


Fig. 1- Initial screen gives the introduction of the application and prompts the user to input the data file along with some input options

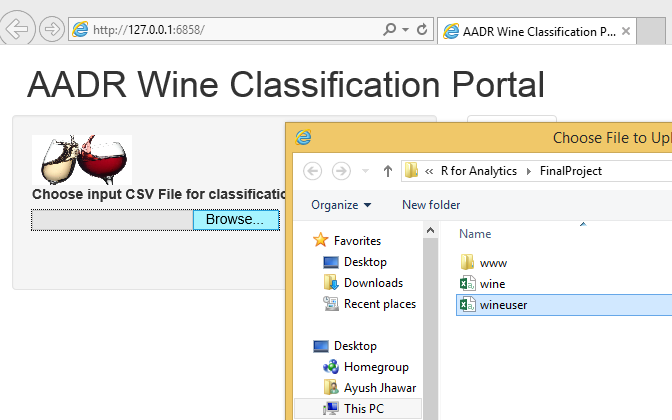


Fig. 2 – The screen shows the file browsing from user machine

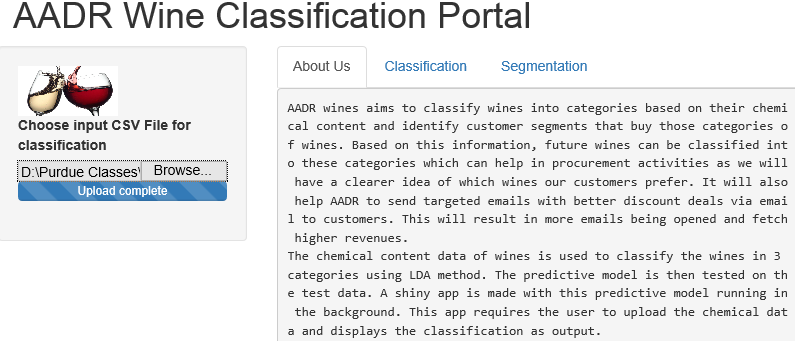


Fig. 3 – The screen shows that the file has been successfully uploaded

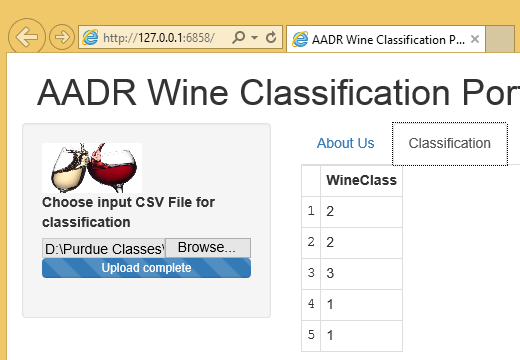


Fig. 4 – The screen shows wine classification in the Classification tab based on the entries in the input file

****

Fig. 5 – The screen shows market segmentation as per wine classification

**Conclusions**

The analysis on the available data using linear discriminant analysis technique gives a predictive model with 98% accuracy. A shiny app with this model in the background classifies new data entered by the user. It helps AADR to market their offerings to a more specific audience which helps them generate higher revenues.

## References (if any)

* <http://math.arizona.edu/~hzhang/math574m/LogitOrLDA.pdf>
* <http://stats.stackexchange.com/questions/95247/logistic-regression-vs-lda-as-two-class-classifiers>
* <http://archive.ics.uci.edu/ml/datasets/Wine>
* <https://mycourses.purdue.edu/bbcswebdav/pid-7737076-dt-content-rid-29985830_1/courses/wl_18128.201710/Beginning%20Shiny.pdf>
* <https://www.rstudio.com/resources/webinars/how-to-start-with-shiny-part-1/>
* <https://www.rstudio.com/resources/webinars/how-to-start-with-shiny-part-2/>
* <https://www.rstudio.com/resources/webinars/how-to-start-with-shiny-part-3/>
* [bit.ly/shiny-quickstart-1](http://www.bit.ly/shiny-quickstart-1)

**Code block:**

**#Modeling Starts**

library(shiny)

#Loading the data

install.packages(caret)

#Reading the data

winedata <- read.csv("D:/Purdue Classes/R for Analytics/FinalProject/wine.csv", header = TRUE)

str(winedata)

names(winedata)

class(winedata)

winedata

# converting class into factor variable

winedata$Class <- as.factor(winedata$Class)

class(winedata$Class)

str(winedata)

library(caret)

inTrain <- createDataPartition(winedata$Class,

p=.70,

list=F)

train <- winedata[inTrain,]

test <- winedata[-inTrain,]

nrow(train)

nrow(test)

ldafit <- train(Class ~.,

data=train,

method="lda")

# make predictions

ldaClasses <- predict(ldafit, newdata=test)

summary(ldaClasses)

# evaluate model

a<- confusionMatrix(data=ldaClasses, test$Class)

a

**#Modeling ends**

**#UI Starts**

ui <- fluidPage(

headerPanel("AADR Wine Classification Portal"),

sidebarLayout(

sidebarPanel(

img(src = "wine.jpg", height = 50, width = 100),

fileInput('file1', 'Choose input CSV File for classification',

accept=c('text/csv',

'text/comma-separated-values,text/plain',

'.csv'))

),

mainPanel(

tabsetPanel(

tabPanel("About Us",verbatimTextOutput("about")),

tabPanel("Classification",tableOutput("table1")),

tabPanel("Segmentation",tableOutput("ms"))

)

)

)

)

**#UI Ends**

**#Server Starts**

server <- function(input,output){

dataInput = reactive({

in.file = input$file1

if (is.null(in.file))

return(NULL)

read.csv(in.file$datapath)

})

clusters <- reactive({

df <- dataInput()

WineClass <- predict(ldafit, newdata = df)

dframe <- as.data.frame(WineClass)

})

output$about <- renderText({

HTML(paste0("AADR wines aims to classify wines into categories based on their chemical content and identify customer segments that buy those categories of wines. Based on this information, future wines can be classified into these categories which can help in procurement activities as we will have a clearer idea of which wines our customers prefer. It will also help AADR to send targeted emails with better discount deals via email to customers. This will result in more emails being opened and fetch higher revenues.

The chemical content data of wines is used to classify the wines in 3 categories using LDA method. The predictive model is then tested on the test data. A shiny app is made with this predictive model running in the background. This app requires the user to upload the chemical data and displays the classification as output."))

})

output$ms <- renderText({

HTML(paste0("<br/><b>","Target Market according to class","</b><br/><br/>",

"1: Preffered by the states of Indiana and North Carolina","<br/>",

"2: Preffered by the states of Colorado and Texas","<br/>",

"3: Preffered by the states of Ohio and Nebraska"))

})

output$table1 <- renderTable({

toprint = clusters()

})

}

**# Server Ends**

**#Application Starts**

shinyApp(ui=ui, server=server)

**#Application Ends**