ASSIGNMENT 4

Heart_disease.CSV dataset contains medical information used for predicting heart disease. It includes various attributes such as age, sex, chest pain type, blood pressure, cholesterol levels, fasting blood sugar, resting ECG results, MAX_HeartRate, and exercise. Our objective is to predict whether a person experiencing chest pain will develop heart disease or not. To achieve this, you will create a dummy variable called "Target" that will be "Yes" if the "MAX_HeartRate" exceeds 170 and "No" if it is 170 or below. Additionally, you will create another dummy variable named "BP_New," which will be "Yes" if the blood pressure is above 120 and "No" if it is 120 or below.
Relevant libraries: caret, e1071, dplyr, tinytex.

Answer the following three questions

Q1) Prediction Based on Initial Information: [30 Points] If a person with chest pain presents without any further information, what prediction should we make regarding heart disease (Yes or No)? Please explain your reasoning.

```
Hint:  Target_table <- table(Heart_disease$Target)
```

Q2) Analysis of the First 30 Records: [3*20=60 Points] Select the first 30 records in the dataset and focus on the "Target" response variable and the two predictors: blood pressure and chest pain type. Create a pivot table that examines heart disease as a function of these two predictors for these 30 records, incorporating all three variables as rows and columns.

a. **Compute Bayes Conditional Probabilities**: Calculate the exact Bayes conditional probabilities of the "Target (Target=yes)" variable given the four possible combinations of the predictors.

```
Hint: Partial code view:
Heart_disease30 <-
Heart_disease[1:30,c("Target","BP_New","chest_pain_type")]
Object1 <- ftable(Heart_disease30)
…
#Four Outcomes P1,…,P4
P1 = Object1[3,1] / Object2[1,1] # Target=yes, BP_New=No &
chest_pain_type=0
```

b. **Classification of Accidents**: Classify the 30 records using these probabilities with a cutoff of 0.5.

```
Hint: Partial code view:
Probability_Target <- rep(0,30)
for (i in 1:30) {…}

Heart_disease30$Probability_Target <- Probability_Target
Heart_disease30$Pred_Probability <-
ifelse(Heart_disease30$Probability_Target > 0.5,  "Yes", "No")
```

c. **Manual Calculation of Naive Bayes Probability**: Manually compute the naive Bayes conditional probability of an injury given that "BP_New" is "Yes" and "chest_pain_type" is 1.

Q3) Full Dataset Analysis: [50 Points] Now, use the complete dataset. Partition the data into training (60%) and validation (40%) sets. Run the Naive Bayes classifier on the entire training set using the relevant predictors, with "Target" as the response variable. Note that all predictors are categorical. Present the confusion matrix.

Hint: Partial code view:

```
nb <- naiveBayes(Target ~chest_pain_type + BP_New, data =
Heart_disease30)

#Splitting the data into 60% training and 40% validation.
train.index <-
sample(row.names(Heart_disease),0.6*dim(Heart_disease)[1])
valid.index <- setdiff(row.names(Heart_disease),train.index)
train.df <- Heart_disease[train.index,]
valid.df <- Heart_disease[valid.index,]

train.norm.df <- train.df[,-9]
…
nb_model <- naiveBayes(…)
train_pred <- predict(…)
conf_matrix1 <- confusionMatrix(…)
```

Note that there is no unique way to develop your code; provided partial codes (pseudocodes) just serve you as hints and you don't need to use them.

- Submit your report in PDF format using R-Markdown. It must include your interpretation/explanations, codes, and outputs.