Supervised Learning

DM 101: Laboratory Exam #3

Ram Reniel Canido, Jeshua Lexis Labio, C.J. Odato, Jan Iris Oiga, Virgilio Salcedo II

2025-03-11

Load Necessary Packages

```
library(readxl)
library(ggplot2)
library(caret)
library(rpart)
library(rpart.plot)
library(lattice)
library(zoo)
library(moments)
```

Load Dataset

```
accidentData <- read_excel("C:/Users/Micah/Desktop/DM101 LabExam3/accident.xlsx")
forPrediction <- read_excel("C:/Users/Micah/Desktop/DM101 LabExam3/for_prediction.xlsx")
```

Check for Missing Values

```
colSums(is.na(accidentData))
```

```
##             Age          Gender Speed_of_Impact     Helmet_Used   Seatbelt_Used
##               0               1               3               0               0
##        Survived
##               0
```

Impute missing values for Gender and Speed_of_Impact

```
#Since Gender is categorical, the best approach is to impute using the mode (most
#frequent value).
get_mode <- function(x) {
  unique_x <- unique(x[!is.na(x)])
  unique_x[which.max(tabulate(match(x, unique_x)))]
```

```
}
accidentData$Gender[is.na(accidentData$Gender)] <- get_mode(accidentData$Gender)

#Impute Speed_of_Impact Based on Skewness
skew_value <- skewness(accidentData$Speed_of_Impact, na.rm = TRUE)
cat("\nSkewness of Speed_of_Impact:", skew_value, "\n")
```

```
##
## Skewness of Speed_of_Impact: 0.05122017
```

```
#Since it is greater than 0.05, we will use median
accidentData$Speed_of_Impact[is.na(accidentData$Speed_of_Impact)] <-
  median(accidentData$Speed_of_Impact, na.rm = TRUE)
```

Standardize Categorical Variables

```
#Convert as Factor
accidentData$Gender <- as.factor(accidentData$Gender)
accidentData$Helmet_Used <- as.factor(accidentData$Helmet_Used)
accidentData$Seatbelt_Used <- as.factor(accidentData$Seatbelt_Used)
accidentData$Survived <- as.factor(accidentData$Survived)
```

Normalize Numerical Variables (Speed_of_Impact)

```
# Convert Speed_of_Impact to numeric before transformation
accidentData$Speed_of_Impact <- as.numeric(scale(accidentData$Speed_of_Impact))

# Split data into training and testing sets first
set.seed(43)
trainIndex <- createDataPartition(accidentData$Survived, p = 0.8, list = FALSE)
train_data <- accidentData[trainIndex, ]
test_data <- accidentData[-trainIndex, ]
```

1. Train Models

```
# Define training control for cross-validation
trctrl <- trainControl(method = "cv", number = 10)

# (1) Logistic Regression
logModel <- train(Survived ~ Gender + Speed_of_Impact + Helmet_Used + Seatbelt_Used,
                  data = train_data,
                  method = "glm",
                  family = "binomial",
                  trControl = trctrl,
                  preProcess = c("center", "scale"),
                  metric = "Accuracy")
```

```r
# (2) k-Nearest Neighbors (k-NN)
knnModel <- train(Survived ~ Gender + Speed_of_Impact + Helmet_Used + Seatbelt_Used,
                  data = train_data,
                  method = "knn",
                  trControl = trctrl,
                  preProcess = c("center", "scale"),
                  tuneLength = 10)

# (3) Support Vector Machines (SVM)
svmModel <- train(Survived ~ Gender + Speed_of_Impact + Helmet_Used + Seatbelt_Used,
                  data = train_data,
                  method = "svmRadial",
                  trControl = trctrl,
                  preProcess = c("center", "scale"),
                  tuneLength = 10)

# (4) Naïve Bayes
nbModel <- train(Survived ~ Gender + Speed_of_Impact + Helmet_Used + Seatbelt_Used,
                 data = train_data,
                 method = "naive_bayes",
                 trControl = trctrl,
                 preProcess = c("center", "scale"))

# (5) Decision Tree
dtModel <- train(Survived ~ Gender + Speed_of_Impact + Helmet_Used + Seatbelt_Used,
                 data = train_data,
                 method = "rpart",
                 trControl = trctrl,
                 preProcess = c("center", "scale"))
```

2. Model Performance Evaluation

```r
logPred <- predict(logModel, test_data)
knnPred <- predict(knnModel, test_data)
svmPred <- predict(svmModel, test_data)
nbPred <- predict(nbModel, test_data)
dtPred <- predict(dtModel, test_data)


logCm <- confusionMatrix(logPred, test_data$Survived)
knnCm <- confusionMatrix(knnPred, test_data$Survived)
svmCm <- confusionMatrix(svmPred, test_data$Survived)
nbCm <- confusionMatrix(nbPred, test_data$Survived)
dtCm <- confusionMatrix(dtPred, test_data$Survived)

#Performance metrics
print(logCm)
```

```
## Confusion Matrix and Statistics
##
##           Reference
```

```
## Prediction  0  1
##          0 11 13
##          1  8  7
##
##                Accuracy : 0.4615
##                  95% CI : (0.3009, 0.6282)
##     No Information Rate : 0.5128
##     P-Value [Acc > NIR] : 0.7884
##
##                   Kappa : -0.0706
##
##  Mcnemar's Test P-Value : 0.3827
##
##             Sensitivity : 0.5789
##             Specificity : 0.3500
##          Pos Pred Value : 0.4583
##          Neg Pred Value : 0.4667
##              Prevalence : 0.4872
##          Detection Rate : 0.2821
##    Detection Prevalence : 0.6154
##       Balanced Accuracy : 0.4645
##
##        'Positive' Class : 0
##
```

```r
print(knnCm)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 15 13
##          1  4  7
##
##                Accuracy : 0.5641
##                  95% CI : (0.3962, 0.7219)
##     No Information Rate : 0.5128
##     P-Value [Acc > NIR] : 0.31610
##
##                   Kappa : 0.1378
##
##  Mcnemar's Test P-Value : 0.05235
##
##             Sensitivity : 0.7895
##             Specificity : 0.3500
##          Pos Pred Value : 0.5357
##          Neg Pred Value : 0.6364
##              Prevalence : 0.4872
##          Detection Rate : 0.3846
##    Detection Prevalence : 0.7179
##       Balanced Accuracy : 0.5697
##
##        'Positive' Class : 0
##
```

```r
print(svmCm)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 14 16
##          1  5  4
##
##                Accuracy : 0.4615
##                  95% CI : (0.3009, 0.6282)
##     No Information Rate : 0.5128
##     P-Value [Acc > NIR] : 0.7884
##
##                   Kappa : -0.0623
##
##  Mcnemar's Test P-Value : 0.0291
##
##             Sensitivity : 0.7368
##             Specificity : 0.2000
##          Pos Pred Value : 0.4667
##          Neg Pred Value : 0.4444
##              Prevalence : 0.4872
##          Detection Rate : 0.3590
##    Detection Prevalence : 0.7692
##       Balanced Accuracy : 0.4684
##
##        'Positive' Class : 0
##
```

```r
print(nbCm)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 11 13
##          1  8  7
##
##                Accuracy : 0.4615
##                  95% CI : (0.3009, 0.6282)
##     No Information Rate : 0.5128
##     P-Value [Acc > NIR] : 0.7884
##
##                   Kappa : -0.0706
##
##  Mcnemar's Test P-Value : 0.3827
##
##             Sensitivity : 0.5789
##             Specificity : 0.3500
##          Pos Pred Value : 0.4583
##          Neg Pred Value : 0.4667
```

```
##               Prevalence : 0.4872
##           Detection Rate : 0.2821
##     Detection Prevalence : 0.6154
##        Balanced Accuracy : 0.4645
##
##         'Positive' Class : 0
##
```

```r
print(dtCm)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 15 15
##          1  4  5
##
##                Accuracy : 0.5128
##                  95% CI : (0.3478, 0.6758)
##     No Information Rate : 0.5128
##     P-Value [Acc > NIR] : 0.56403
##
##                   Kappa : 0.0389
##
##  Mcnemar's Test P-Value : 0.02178
##
##             Sensitivity : 0.7895
##             Specificity : 0.2500
##          Pos Pred Value : 0.5000
##          Neg Pred Value : 0.5556
##              Prevalence : 0.4872
##          Detection Rate : 0.3846
##    Detection Prevalence : 0.7692
##       Balanced Accuracy : 0.5197
##
##        'Positive' Class : 0
##
```

The performance of five classification models—Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT)—was evaluated using confusion matrices and various statistical metrics.

Key Findings:

Accuracy & Balanced Accuracy:

The highest accuracy was achieved by KNN (56.41%), followed by Decision Tree (51.28%). Logistic Regression, SVM, and Naïve Bayes performed similarly with 46.15% accuracy. Balanced accuracy ranged between 46.45% (Logistic Regression, NB) to 56.97% (KNN).

Sensitivity & Specificity:

Sensitivity (ability to detect positive cases) was highest in KNN (78.95%) and Decision Tree (78.95%). Specificity (ability to detect negative cases) was very low across all models, especially in SVM (20%) and Decision Tree (25%).

Kappa & McNemar's Test:

Kappa scores were close to 0 or negative, indicating poor agreement beyond chance. McNemar's test showed statistical significance for SVM and Decision Tree, suggesting imbalanced misclassification rates.

Predictive Values:

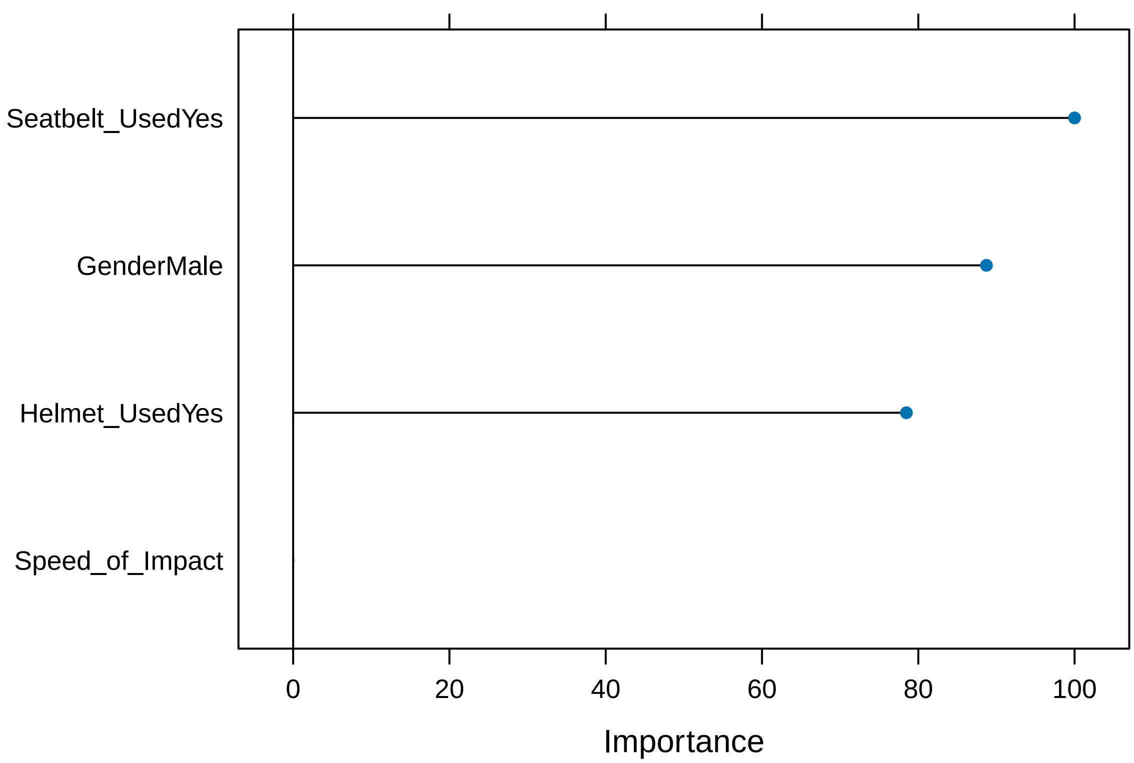
Positive Predictive Value (PPV) and Negative Predictive Value (NPV) varied slightly, with KNN showing a higher NPV (63.64%) compared to others.

Are All Models Useful for Classification?

No, not all models are equally useful in this case. The overall accuracy and balanced accuracy suggest that these models struggle with classification, likely due to data imbalance or feature limitations. KNN performed the best, with relatively higher accuracy, sensitivity, and balanced accuracy, making it a better choice. Logistic Regression and Naïve Bayes performed similarly, but their accuracy was below 50%, making them unreliable. SVM showed the weakest specificity, meaning it struggles to correctly classify negative cases. Decision Tree had reasonable sensitivity but poor specificity, making it unreliable for balanced classification. For practical classification, KNN or Decision Tree may be considered, but improvements such as feature selection, hyperparameter tuning, or handling class imbalance (e.g., resampling techniques) should be explored for better performance.

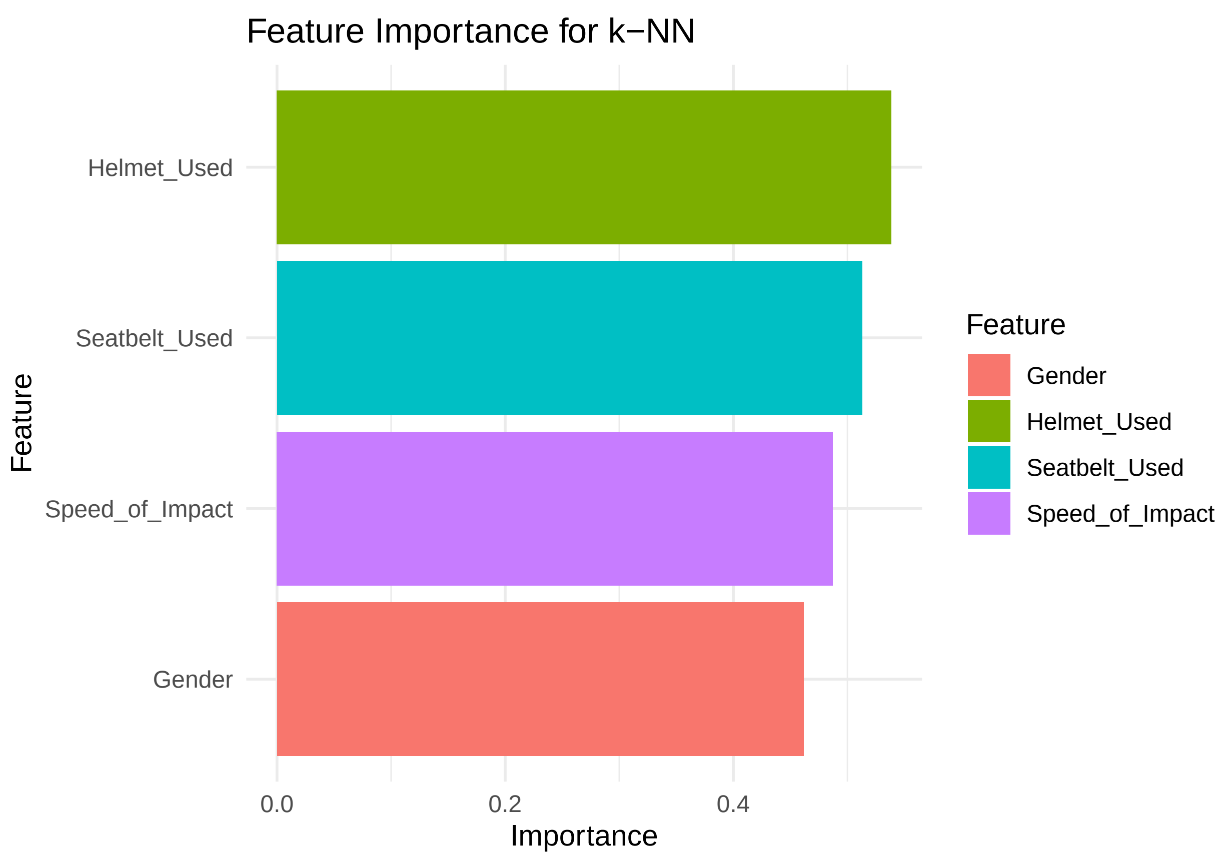
3. Important features or predictors for each model

```
logImp <- varImp(logModel)
plot(logImp)
```



The Logistic Regression model suggests that seatbelt usage is the most critical factor for survival, indicating that wearing a seatbelt significantly improves the chances of surviving an accident. Gender (male) and helmet usage also play important roles, with helmet use contributing positively to survival in motorcycle-related incidents. Speed of impact appears to have the least influence in this model, which could mean that protective measures (seatbelts and helmets) mitigate the effects of speed to some extent. Overall, proper safety measures, especially seatbelt usage, are important in enhancing survival rates in accidents.

```
knn_imp <- data.frame(Feature = character(), Importance = numeric(), stringsAsFactors = FALSE)
features <- c("Gender", "Speed_of_Impact", "Helmet_Used", "Seatbelt_Used")

for (feature in features) {
  temp_model <- train(as.formula(paste("Survived ~", paste(setdiff(features, feature),
                                                           collapse = " + "))),
                      data = train_data, method = "knn",
                      trControl = trainControl(method = "cv", number = 10),
                      preProcess = c("center", "scale"), tuneLength = 10)
  temp_predictions <- predict(temp_model, test_data, type = "raw")
  importance_score <- mean(temp_predictions == test_data$Survived)
  knn_imp <- rbind(knn_imp, data.frame(Feature = feature, Importance = importance_score))
}

ggplot(knn_imp, aes(x = reorder(Feature, Importance), y = Importance, fill = Feature)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Feature Importance for k-NN", x = "Feature", y = "Importance") +
  theme_minimal()
```

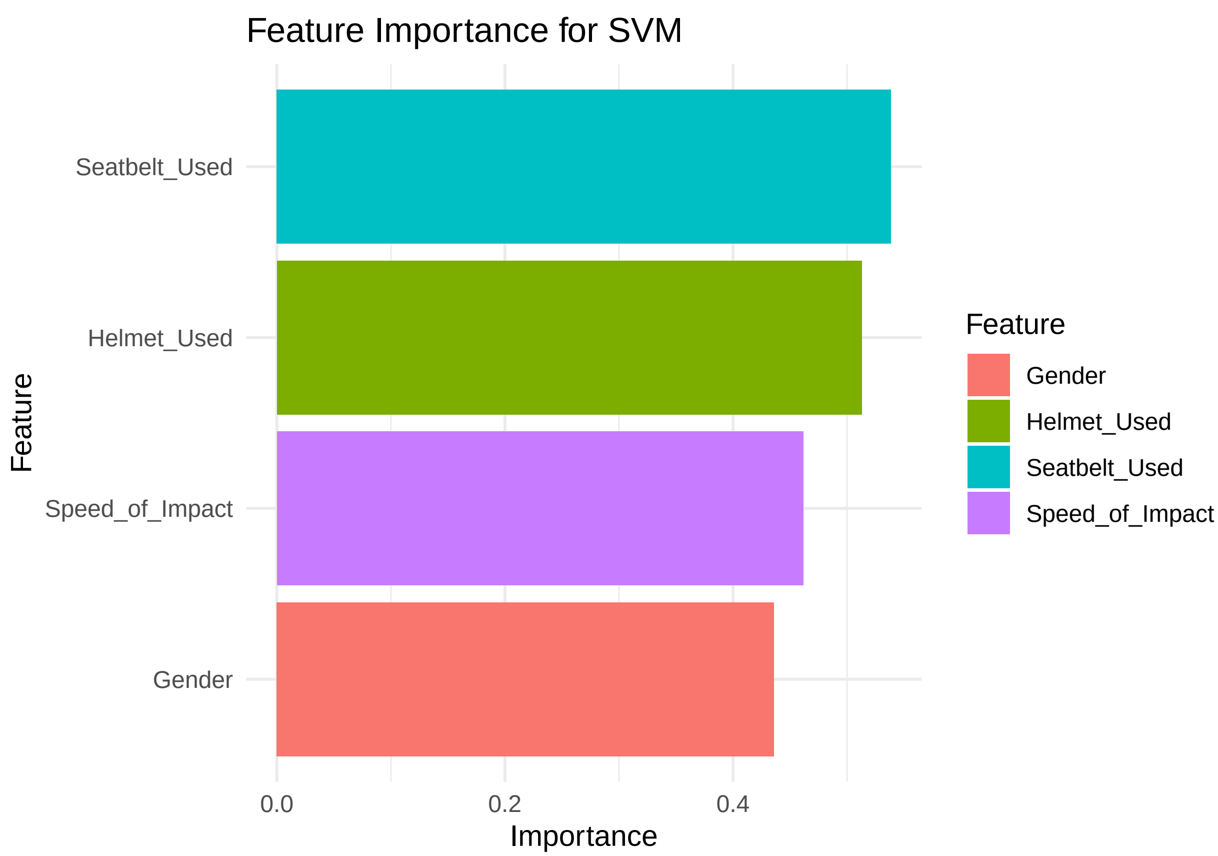


Helmet_Used is the most significant factor, emphasizing the strong role of helmet protection in survival.

Seatbelt_Used also plays a crucial role, reinforcing the importance of safety gear. Speed_of_Impact is important but slightly less than helmet and seatbelt use, suggesting that while speed affects survival, protective measures can mitigate risk. Gender has the lowest importance, indicating that it does not substantially influence survival outcomes.

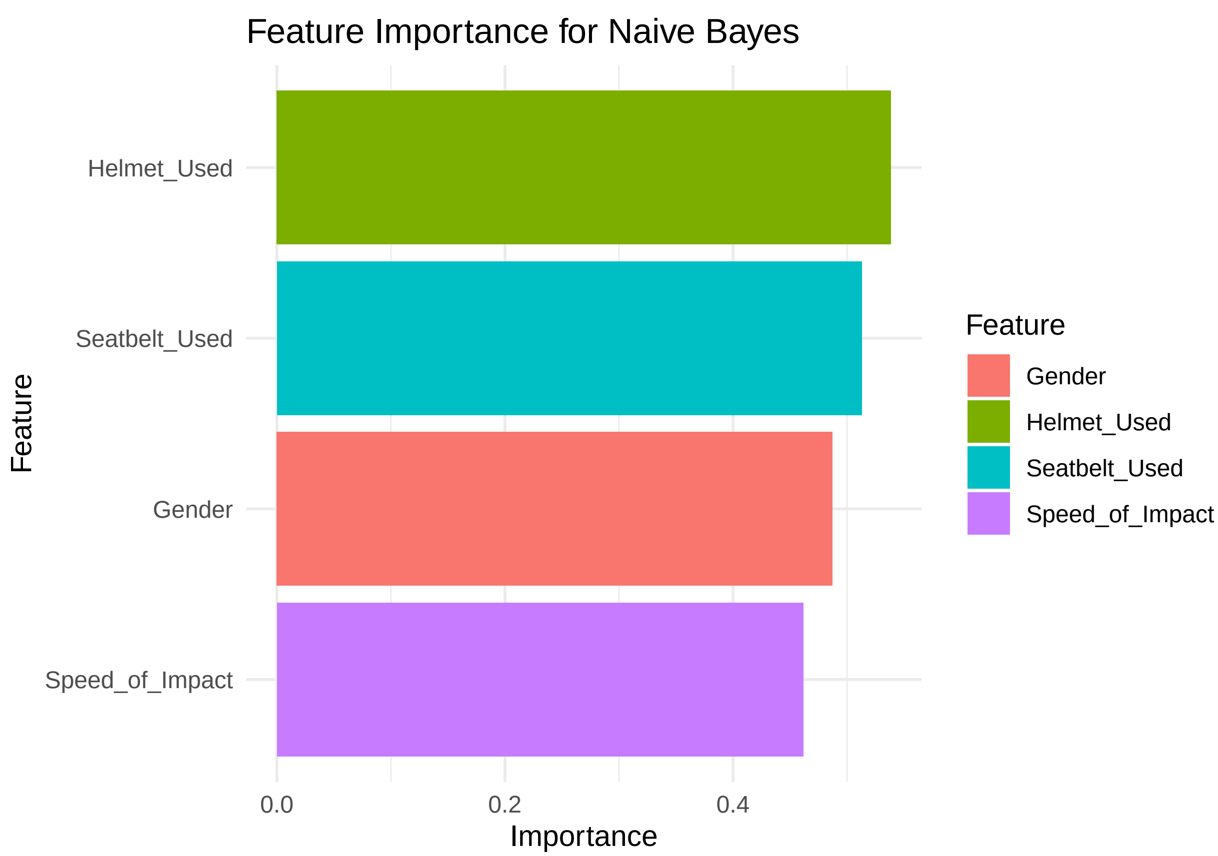
This model suggests that helmet and seatbelt use are the best predictors of survival, highlighting the importance of protective gear over other factors.

```r
svm_imp <- data.frame(Feature = character(), Importance = numeric(), stringsAsFactors = FALSE)

for (feature in features) {
  temp_data <- test_data
  temp_data[[feature]] <- sample(temp_data[[feature]])
  temp_predictions <- predict(svmModel, temp_data, type = "raw")
  importance_score <- mean(temp_predictions == test_data$Survived)
  svm_imp <- rbind(svm_imp, data.frame(Feature = feature, Importance = importance_score))
}

ggplot(svm_imp, aes(x = reorder(Feature, Importance), y = Importance, fill = Feature)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Feature Importance for SVM", x = "Feature", y = "Importance") +
  theme_minimal()
```



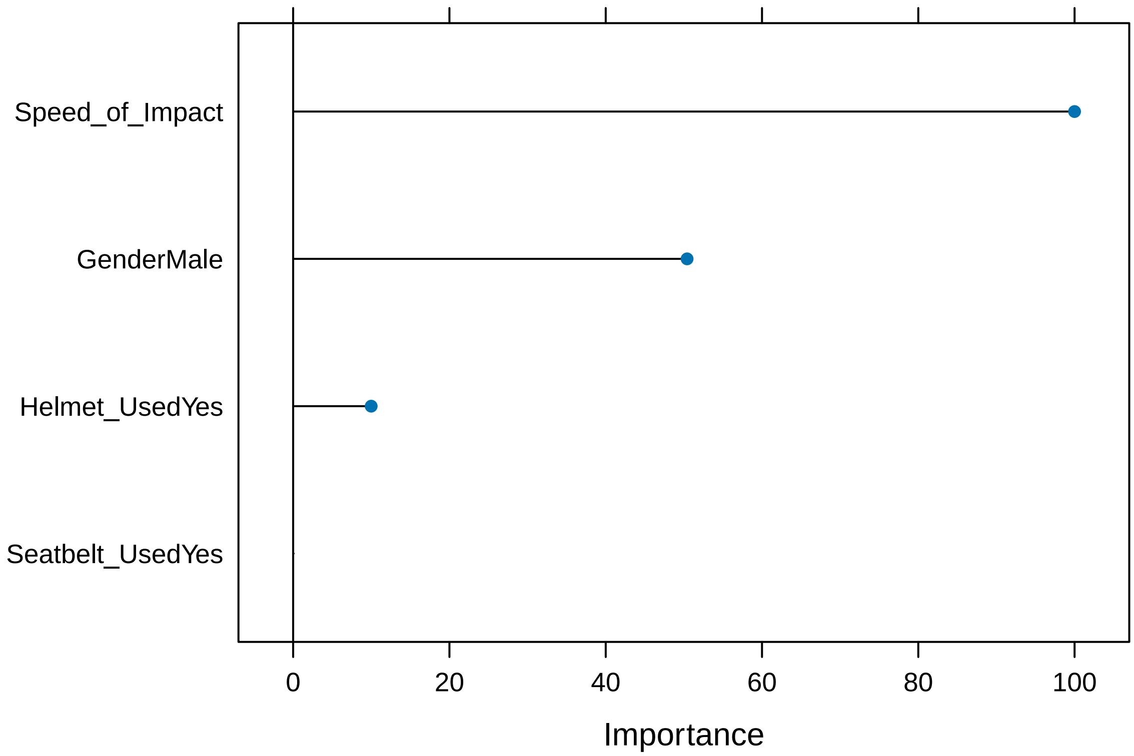
The SVM model highlights that seatbelt usage is the most crucial factor influencing survival, indicating that individuals who wear seatbelts have significantly higher chances of surviving an accident. Helmet usage follows closely, emphasizing its critical role in preventing fatal head injuries, especially for motorcyclists and those in open vehicles. Speed of impact also contributes to survival probability, as higher speeds generally result in more severe injuries, though its impact is slightly less than safety measures like seatbelt and helmet use. Gender, while having some influence, is the least important factor, suggesting that biological differences play a minimal role in survival compared to safety precautions and accident severity. Overall, this analysis reinforces the importance of wearing seatbelts and helmets, as they are the strongest predictors of survival, and highlights the need for strict enforcement of road safety measures.

```
nbayes_imp <- data.frame(Feature = character(), Importance = numeric(), stringsAsFactors = FALSE)

for (feature in features) {
  temp_data <- test_data
  temp_data[[feature]] <- sample(temp_data[[feature]])
  temp_predictions <- predict(nbModel, temp_data, type = "raw")
  importance_score <- mean(temp_predictions == test_data$Survived)
  nbayes_imp <- rbind(nbayes_imp, data.frame(Feature = feature, Importance = importance_score))
}

ggplot(nbayes_imp, aes(x = reorder(Feature, Importance), y = Importance, fill = Feature)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Feature Importance for Naive Bayes", x = "Feature", y = "Importance") +
  theme_minimal()
```



The Naïve Bayes model indicates that helmet usage is the most significant factor affecting survival, emphasizing that wearing a helmet greatly increases the likelihood of surviving an accident. Seatbelt usage is also highly influential, reinforcing the importance of protective measures in minimizing fatal injuries. Interestingly, gender plays a relatively larger role in this model compared to others, suggesting that biological or behavioral differences might slightly influence survival outcomes. Speed of impact has the lowest importance, implying that while crash severity is relevant, safety equipment like helmets and seatbelts are stronger predictors of survival. Overall, this analysis highlights the critical role of safety gear in accident survival, reinforcing the need for consistent use of helmets and seatbelts to improve road safety outcomes.

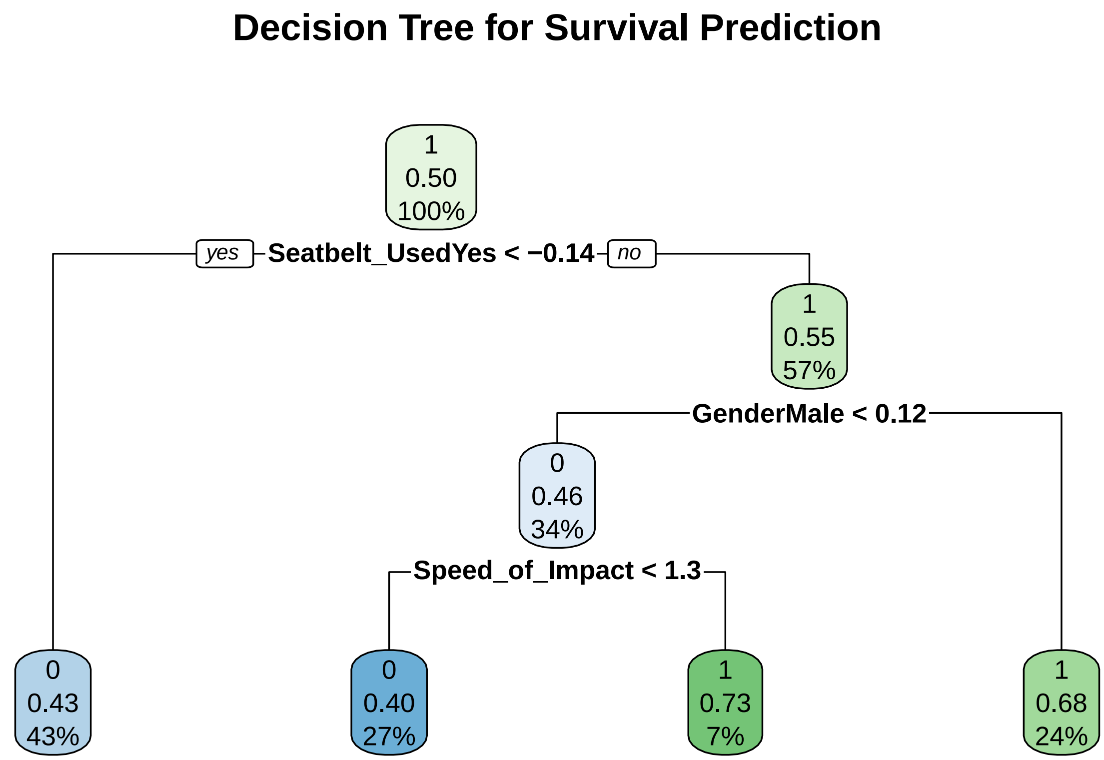
```
dtImp <- varImp(dtModel)
plot(dtImp)
```



The decision tree model suggests that survival in accidents is most strongly influenced by the speed of impact, with higher speeds reducing the likelihood of survival. Gender (male) also plays a significant role, potentially due to behavioral or exposure differences. While helmet and seatbelt usage contribute to survival, their influence is less pronounced compared to speed. This highlights that while protective measures are important, reducing speed remains the most critical factor in improving survival rates in accidents.

4. Decision tree diagram

```
rpart.plot(dtModel$finalModel, main = "Decision Tree for Survival Prediction")
```



The decision tree for survival prediction highlights that seatbelt usage is the most critical factor influencing survival rates. Individuals who wear a seatbelt have a significantly higher chance of survival, with probabilities reaching 68–73%, even at varying speeds of impact. Conversely, those who do not use a seatbelt face a much higher risk of fatality, with survival probabilities dropping below 50%. Gender also plays a role, as females appear to have slightly lower survival rates when not wearing a seatbelt. Additionally, speed of impact influences survival, with lower speeds being associated with reduced survival probabilities in some cases. Overall, the model strongly emphasizes the life-saving importance of wearing a seatbelt, regardless of speed or gender.

5. Prediction using the Best Model

```r
#Determining the best model
# Extract accuracy, sensitivity, specificity, and Kappa
model_results <- data.frame(
  Model = c("Logistic Regression", "k-NN", "SVM", "Naive Bayes", "Decision Tree"),
  Accuracy = c(logCm$overall["Accuracy"],
               knnCm$overall["Accuracy"],
               svmCm$overall["Accuracy"],
               nbCm$overall["Accuracy"],
               dtCm$overall["Accuracy"]),
  Sensitivity = c(logCm$byClass["Sensitivity"],
                  knnCm$byClass["Sensitivity"],
                  svmCm$byClass["Sensitivity"],
                  nbCm$byClass["Sensitivity"],
                  dtCm$byClass["Sensitivity"]),
```

```r
  Specificity = c(logCm$byClass["Specificity"],
                  knnCm$byClass["Specificity"],
                  svmCm$byClass["Specificity"],
                  nbCm$byClass["Specificity"],
                  dtCm$byClass["Specificity"]),
  Kappa = c(logCm$overall["Kappa"],
            knnCm$overall["Kappa"],
            svmCm$overall["Kappa"],
            nbCm$overall["Kappa"],
            dtCm$overall["Kappa"])
)

print(model_results)
```

```
##                  Model  Accuracy Sensitivity Specificity       Kappa
## 1 Logistic Regression 0.4615385   0.5789474        0.35 -0.07058824
## 2                k-NN 0.5641026   0.7894737        0.35  0.13784135
## 3                 SVM 0.4615385   0.7368421        0.20 -0.06225681
## 4         Naive Bayes 0.4615385   0.5789474        0.35 -0.07058824
## 5       Decision Tree 0.5128205   0.7894737        0.25  0.03891051
```

Our group decided to use k-Nearest Neighbors (k-NN) model. This model has the highest accuracy (0.5641), meaning it makes the most correct predictions overall.It also got the highest sensitivity (0.7895), which indicates it performs well in identifying positive cases. A moderate kappa score (0.1378), which is better than most other models in agreement beyond chance.

Standardize Categorical Variables for Prediction

```r
forPrediction$Gender <- ifelse(forPrediction$Gender == "F", "Female",
                ifelse(forPrediction$Gender == "M", "Male", forPrediction$Gender))
forPrediction$Helmet_Used <- ifelse(forPrediction$Helmet_Used == "Y", "Yes",
        ifelse(forPrediction$Helmet_Used == "N", "No", forPrediction$Helmet_Used))
forPrediction$Seatbelt_Used <- ifelse(forPrediction$Seatbelt_Used == "Y", "Yes",
    ifelse(forPrediction$Seatbelt_Used == "N", "No", forPrediction$Seatbelt_Used))

#Convert as Factor
forPrediction$Gender <- as.factor(forPrediction$Gender)
forPrediction$Helmet_Used <- as.factor(forPrediction$Helmet_Used)
forPrediction$Seatbelt_Used <- as.factor(forPrediction$Seatbelt_Used)
forPrediction$Speed_of_Impact <- as.numeric(scale(forPrediction$Speed_of_Impact))

#Predict survival on new dataset
forPrediction$Survival_Rate <- predict(knnModel, newdata = forPrediction, type = "raw")
print(head(forPrediction[, c("Gender", "Speed_of_Impact", "Helmet_Used",
                              "Seatbelt_Used", "Survival_Rate")]))
```

```
## # A tibble: 6 x 5
##   Gender Speed_of_Impact Helmet_Used Seatbelt_Used Survival_Rate
##   <fct>            <dbl> <fct>       <fct>         <fct>
## 1 Female            1.38 No          Yes           1
```

```
## 2 Female          -1.72  No          Yes          0
## 3 Male             0.442 No          Yes          1
## 4 Male             0.586 Yes         Yes          1
## 5 Female           1.13  No          No           0
## 6 Male             1.16  Yes         No           0
```

```r
print(forPrediction$Survival_Rate)
```

```
##  [1] 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 0 0 1 1 1 1 0 1 0
## [39] 1 1 0 0 1 1 0 1 1 1 0 0 0 1 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 1 0 1 0 1 0 1 1
## [77] 0 0 1 1 1 1 1 0 0 0 0 1 0 0 1 0 1 1 0 1 0 1 1
## Levels: 0 1
```

6. Other feature variables or predictors

1. Type of Vehicle Involved

The survival rate differs depending on whether the individual was in a motorcycle, car, truck, or bicycle. Motorcyclists, for example, have a much higher fatality risk due to a lack of protective structure. Research by Savolainen et al. (2011) shows that motorcyclists are significantly more vulnerable to fatal injuries compared to car passengers.

2. Road Condition at the Time of the Accident

Slippery or poorly maintained roads increase accident severity, leading to more fatal injuries. Poor road conditions may also reduce braking efficiency and vehicle stability. A study in the Journal of Safety Research (El-Basyouny & Sayed, 2013) found that wet or icy roads significantly increase the probability of severe injuries.

3. Time of the Accident (Day/Night)

Nighttime accidents tend to be more severe due to reduced visibility, driver fatigue, and a higher likelihood of impaired driving. Clarke et al. (2006) found that nighttime accidents have higher fatality rates, especially due to drunk driving and poor lighting conditions.

4. Alcohol Influence (Yes/No or Blood Alcohol Level)

Alcohol impairs reaction time, judgment, and coordination, increasing the likelihood of severe accidents and fatalities. The Centers for Disease Control and Prevention (CDC) reports that nearly 30% of traffic fatalities in the U.S. are caused by alcohol-impaired driving.

5. Speed Limit of the Road

Accidents on high-speed roads (e.g., highways) tend to be more severe than those in low-speed zones. A higher speed increases impact force, reducing survival chances. Rosén et al. (2011) found that the probability of fatality doubles when speed increases from 50 km/h to 70 km/h.

6. Seat Position in the Vehicle (Driver, Front Passenger, Back Passenger)

The driver and front-seat passengers are at a higher risk of fatal injuries compared to rear-seat passengers, who are better protected. Research in the Journal of Traffic Injury Prevention (2009) states that rear-seat passengers have a 25-30% higher survival rate than those in the front seats.

TEAM'S CONTRIBUTION

Canido, Ram Reniel

Canido took responsibility for coding questions 3 and 5, ensuring that the necessary functions and algorithms were correctly implemented. Beyond coding, Canido guides the team, giving clear instructions on what needed to be done at each stage of the coding process, and ensuring that all team members remained aligned with the project's objectives. Canido's contributions ensured that the technical aspects of the implementation were executed efficiently and accurately..

Labio, Jeshua Lexis

Labio focused on interpreting the key features of each model in question 3, analyzing their significance in influencing survival predictions. Labio also provided a thorough interpretation of the Decision Tree diagram in question 4, breaking down how different factors such as speed of impact, seatbelt usage, and helmet usage influenced classification. Labio's detailed insights helped the team better understand the patterns and relationships in the model, strengthening the analytical portion of the project.

Odato, C.J.

Odato was responsible for coding question 2, ensuring the correct implementation of the statistical models and their computations. In addition to coding, Odato provided an in-depth discussion for question 2, elaborating on key concepts and ensuring that the explanations were well-structured and informative. Odato also took the initiative to refine and polish the R Markdown document, improving its overall clarity, coherence, and formatting. This effort ensured that the final presentation of the project was professional and well-organized.

Oiga, Jan Iris

Oiga handled the coding for questions 1 and 4, ensuring that the statistical models and data analysis were executed correctly. Beyond coding, Oiga played an essential role in refining and improving the final presentation of the R Markdown document, focusing on readability and consistency. Oiga's contributions ensured that the project was not only technically sound but also well-structured and easy to understand.

Salcedo, Virgilio II

Salcedo was responsible for answering question 6, providing well-researched and detailed responses. Additionally, Salcedo conducted background research and collected relevant studies to support the analysis, ensuring that the project was backed by credible information. Salcedo also assisted Labio in analyzing and interpreting the Decision Tree diagram in question 4, contributing valuable insights that strengthened the team's understanding of the model.