Análisis Epidemiológico ENSSEX - Predicción de Depresión

Pablo Cabello, Cristobal Pineda y Felipe Carrasco

2024-12-10

Contents

0. Introducción

```
##
##  Contexto General:
##
## - Los análisis realizados se basan en los datos obtenidos de la Encuesta Nacional de Salud, Sexualida
##
## - La base de datos ENSSEX 2022-2023 es una encuesta poblacional de 20.392 casos considerando un mues
##
##  Selección y descripción de la variable dependiente:
##
## 1. Variable Dependiente: 'Depresión'
##
##  Selección y descripción de las variables independientes:
##
## - Variables Categóricas:
##
##   1. Sexo al nacer:
##
##   2. Nivel educativo:
##
##   3. Bienestar emocional (escala del 1 al 7):
##
##   4. Calidad de vida percibida:
##
##   5. Satisfacción aspecto físico:
##
##   6. Consumo de alcohol (en los últimos 30 días):
##
```

```
##   7. Drogas inyectables:
##
## - Variables Numéricas:
##
##   8. Edad (en años):
##
##   9. Peso (en kg):
##
##   10. Talla (en metros):
```

1. Carga de Paquetes y Datos

```r
# Cargar datos
load('20240516_enssex_data.rdata')
```

2. Preparación de Datos

```r
# Crear y limpiar variables
datos <- enssex4 %>%
    mutate(
        # Variable dependiente: convertir a factor y manejar valores faltantes
        depresion = factor(ifelse(!is.na(zap_labels(i_3_p9)) & zap_labels(i_3_p9) %in% c(1,2),
                                  ifelse(zap_labels(i_3_p9) == 1, "Si", "No"), NA)),

        # Variables categóricas con manejo de NA
        sexo_al_nacer = factor(ifelse(!is.na(zap_labels(p1)) & zap_labels(p1) %in% c(1,2),
                                      ifelse(zap_labels(p1) == 1, "Hombre", "Mujer"), NA)),

        nivel_educacional = factor(ifelse(!is.na(zap_labels(p5)),
                                          as.character(as_factor(p5)), NA)),

        bienestar_emocional = factor(ifelse(!is.na(zap_labels(i_2_p9)),
                                            as.character(as_factor(i_2_p9)), NA)),

        calidad_vida_percibida = factor(ifelse(!is.na(zap_labels(p8)),
                                               as.character(as_factor(p8)), NA)),

        satisfaccion_aspecto_fisico = factor(ifelse(!is.na(zap_labels(i_1_p24)),
                                                    as.character(as_factor(i_1_p24)), NA)),

        drogas_inyectables = factor(ifelse(!is.na(zap_labels(i_3_p25)) & zap_labels(i_3_p25) %in% c(1,2
                                           ifelse(zap_labels(i_3_p25) == 1, "Si", "No"), NA)),

        consumo_alcohol = factor(ifelse(!is.na(zap_labels(i_5_p26)),
                                        as.character(as_factor(i_5_p26)), NA)),

        # Variables numéricas con validación
        edad = as.numeric(ifelse(!is.na(zap_labels(p4)) & zap_labels(p4) >= 18 & zap_labels(p4) <= 100,
                                 zap_labels(p4), NA)),
```

```
        peso = as.numeric(ifelse(!is.na(zap_labels(p22)) & zap_labels(p22) >= 30 & zap_labels(p22) <= 20
                                 zap_labels(p22), NA)),

        talla = as.numeric(ifelse(!is.na(zap_labels(p23)) & zap_labels(p23) >= 130 & zap_labels(p23) <=
                                  zap_labels(p23)/100, NA))  # Convertir a metros
    ) %>%
    # Filtrar filas con datos válidos en variables clave
    filter(!is.na(depresion))

# Análisis de datos faltantes
na_summary <- colSums(is.na(datos)) / nrow(datos) * 100
print("Porcentaje de datos faltantes por variable:")
```

```
## [1] "Porcentaje de datos faltantes por variable:"
```

```
print(na_summary[c("depresion", "sexo_al_nacer", "edad", "peso", "talla",
                   "drogas_inyectables", "consumo_alcohol")])
```

```
##          depresion      sexo_al_nacer               edad               peso
##           0.000000           0.000000           0.000000           3.495994
##              talla drogas_inyectables    consumo_alcohol
##           3.532411           1.056082          40.823015
```

3. Análisis Descriptivo

```
## 3. Análisis Descriptivo

# Variables categóricas de interés
vars_cat <- c("depresion", "sexo_al_nacer", "nivel_educacional", "bienestar_emocional",
              "calidad_vida_percibida", "satisfaccion_aspecto_fisico",
              "drogas_inyectables", "consumo_alcohol")

# Variables numéricas de interés
vars_num <- c("edad", "peso", "talla")

# Análisis de frecuencias absolutas y relativas para variables categóricas
for (var in vars_cat) {
    cat("\n--------------------------------\n")
    cat("Análisis de frecuencias para:", var, "\n")

    # Frecuencia absoluta
    tabla_abs <- table(datos[[var]])
    cat("Frecuencias absolutas:\n")
    print(tabla_abs)

    # Frecuencia relativa
    tabla_rel <- prop.table(tabla_abs) * 100
    cat("Frecuencias relativas (%):\n")
    print(round(tabla_rel, 2))
}
```

```
##
## --------------------------------
## Análisis de frecuencias para: depresion
## Frecuencias absolutas:
##
##   No   Si
##  765 1981
## Frecuencias relativas (%):
##
##    No    Si
## 27.86 72.14
##
## --------------------------------
## Análisis de frecuencias para: sexo_al_nacer
## Frecuencias absolutas:
##
## Hombre  Mujer
##    701   2045
## Frecuencias relativas (%):
##
## Hombre  Mujer
##  25.53  74.47
##
## --------------------------------
## Análisis de frecuencias para: nivel_educacional
## Frecuencias absolutas:
##
##                                                Educación Básica
##                                                             671
##                                Educación Especial (Diferencial)
##                                                               1
##                             Educación Media Científico-Humanista
##                                                             882
##                              Educación Media Técnica Profesional
##                                                             173
##                                   Humanidades (Sistema Antiguo)
##                                                             164
##                     Jardín Infantil (medio menor y medio mayor)
##                                                               0
##                                                  Nunca asistió
##                                                              48
##                                              Postgrado Completo
##                                                              10
##                                            Postgrado Incompleto
##                                                               2
##                       Primario o Preparatoria (Sistema antiguo)
##                                                             137
##                 Profesional Completo (carreras de 4 o más años)
##                                                             183
##               Profesional Incompleto (carreras de 4 o más años)
##                                                             149
## Técnica Comercial, Industrial o Normalista (Sistema Antiguo)
##                                                              53
##     Técnico Nivel Superior Completo (carreras de 1 a 3 años)
```

```
##                                                       175
##   Técnico Nivel Superior Incompleto (carreras de 1 a 3 años)
##                                                        98
## Frecuencias relativas (%):
##
##                                            Educación Básica
##                                                       24.44
##                               Educación Especial (Diferencial)
##                                                        0.04
##                           Educación Media Científico-Humanista
##                                                       32.12
##                            Educación Media Técnica Profesional
##                                                        6.30
##                                 Humanidades (Sistema Antiguo)
##                                                        5.97
##                  Jardín Infantil (medio menor y medio mayor)
##                                                        0.00
##                                               Nunca asistió
##                                                        1.75
##                                         Postgrado Completo
##                                                        0.36
##                                       Postgrado Incompleto
##                                                        0.07
##                     Primario o Preparatoria (Sistema antiguo)
##                                                        4.99
##              Profesional Completo (carreras de 4 o más años)
##                                                        6.66
##            Profesional Incompleto (carreras de 4 o más años)
##                                                        5.43
## Técnica Comercial, Industrial o Normalista (Sistema Antiguo)
##                                                        1.93
##     Técnico Nivel Superior Completo (carreras de 1 a 3 años)
##                                                        6.37
##   Técnico Nivel Superior Incompleto (carreras de 1 a 3 años)
##                                                        3.57
##
## ------------------------------
## Análisis de frecuencias para: bienestar_emocional
## Frecuencias absolutas:
##
##  1 Muy mal          2          3          4          5          6 7 Muy bien
##        161        128        232        463        653        620        485
##      Ns-Nr
##          4
## Frecuencias relativas (%):
##
##  1 Muy mal          2          3          4          5          6 7 Muy bien
##       5.86       4.66       8.45      16.86      23.78      22.58      17.66
##      Ns-Nr
##       0.15
##
## ------------------------------
## Análisis de frecuencias para: calidad_vida_percibida
## Frecuencias absolutas:
```

```
##
##            Buena             Mala        Muy buena         Muy mala
##             1269              192              265               42
## Ni buena ni mala      No responde          No sabe
##              968                2                8
## Frecuencias relativas (%):
##
##            Buena             Mala        Muy buena         Muy mala
##            46.21             6.99             9.65             1.53
## Ni buena ni mala      No responde          No sabe
##            35.25             0.07             0.29
##
## --------------------------------
## Análisis de frecuencias para: satisfaccion_aspecto_fisico
## Frecuencias absolutas:
##
##                      De acuerdo                   En desacuerdo
##                            1391                             451
##                  Muy de acuerdo               Muy en desacuerdo
##                             256                             100
## Ni de acuerdo, ni en desacuerdo                              NR
##                             531                               3
##                              NS
##                              14
## Frecuencias relativas (%):
##
##                      De acuerdo                   En desacuerdo
##                           50.66                           16.42
##                  Muy de acuerdo               Muy en desacuerdo
##                            9.32                            3.64
## Ni de acuerdo, ni en desacuerdo                              NR
##                           19.34                            0.11
##                              NS
##                            0.51
##
## --------------------------------
## Análisis de frecuencias para: drogas_inyectables
## Frecuencias absolutas:
##
##   No   Si
## 2713    4
## Frecuencias relativas (%):
##
##    No    Si
## 99.85  0.15
##
## --------------------------------
## Análisis de frecuencias para: consumo_alcohol
## Frecuencias absolutas:
##
##              Durante los últimos 30 días
##                                     1022
##                       Hace más de un año
##                                      279
```

```
## Hace más de un mes, pero menos de un año
##                                       322
##                                     NS/NR
##                                         2
## Frecuencias relativas (%):
##
##              Durante los últimos 30 días
##                                     62.89
##                         Hace más de un año
##                                     17.17
## Hace más de un mes, pero menos de un año
##                                     19.82
##                                     NS/NR
##                                      0.12
```

```
# Análisis descriptivo para variables numéricas
cat("\n--------------------------------\n")
```

```
##
## --------------------------------
```

```
cat("Análisis descriptivo para variables numéricas\n")
```

```
## Análisis descriptivo para variables numéricas
```

```
for (var in vars_num) {
    cat("\n--------------------------------\n")
    cat("Variable:", var, "\n")

    # Resumen de estadísticas descriptivas
    media <- mean(datos[[var]], na.rm = TRUE)
    mediana <- median(datos[[var]], na.rm = TRUE)
    minimo <- min(datos[[var]], na.rm = TRUE)
    maximo <- max(datos[[var]], na.rm = TRUE)

    cat("Media:", round(media, 2), "\n")
    cat("Mediana:", round(mediana, 2), "\n")
    cat("Mínimo:", round(minimo, 2), "\n")
    cat("Máximo:", round(maximo, 2), "\n")
}
```

```
##
## --------------------------------
## Variable: edad
## Media: 50.38
## Mediana: 53
## Mínimo: 18
## Máximo: 100
##
## --------------------------------
## Variable: peso
## Media: 71.93
```

```
## Mediana: 70
## Mínimo: 36
## Máximo: 189
##
## ------------------------------
## Variable: talla
## Media: 1.61
## Mediana: 1.6
## Mínimo: 1.3
## Máximo: 1.96
```

5. Modelado Predictivo

```r
# Cargar librería adicional para imputación
if (!require("glmnet")) install.packages("glmnet")
if (!require("mice")) install.packages("mice")
library(mice)

# Preparar datos para modelado
variables_seleccionadas <- c("depresion", "sexo_al_nacer", "edad", "peso", "talla",
                             "nivel_educacional", "bienestar_emocional", "calidad_vida_percibida",
                             "satisfaccion_aspecto_fisico", "drogas_inyectables", "consumo_alcohol")

datos_modelo <- datos %>%
    dplyr::select(all_of(variables_seleccionadas))

# Verificar el desbalance inicial
print("Distribución inicial de clases:")
```

```
## [1] "Distribución inicial de clases:"
```

```r
print(table(datos_modelo$depresion))
```

```
##
##   No   Si
##  765 1981
```

```r
# Imputación de datos faltantes
# Primero convertimos factores a numéricos para la imputación
datos_numericos <- datos_modelo %>%
    mutate(across(where(is.factor), as.numeric))

# Realizar imputación
imp <- mice(datos_numericos, m=5, maxit=50, method='pmm', seed=123)
```

```
##
##  iter imp variable
##   1   1  peso  talla  drogas_inyectables  consumo_alcohol
##   1   2  peso  talla  drogas_inyectables  consumo_alcohol
##   1   3  peso  talla  drogas_inyectables  consumo_alcohol
```

```
##   1   4  peso  talla  drogas_inyectables  consumo_alcohol
##   1   5  peso  talla  drogas_inyectables  consumo_alcohol
##   2   1  peso  talla  drogas_inyectables  consumo_alcohol
##   2   2  peso  talla  drogas_inyectables  consumo_alcohol
##   2   3  peso  talla  drogas_inyectables  consumo_alcohol
##   2   4  peso  talla  drogas_inyectables  consumo_alcohol
##   2   5  peso  talla  drogas_inyectables  consumo_alcohol
##   3   1  peso  talla  drogas_inyectables  consumo_alcohol
##   3   2  peso  talla  drogas_inyectables  consumo_alcohol
##   3   3  peso  talla  drogas_inyectables  consumo_alcohol
##   3   4  peso  talla  drogas_inyectables  consumo_alcohol
##   3   5  peso  talla  drogas_inyectables  consumo_alcohol
##   4   1  peso  talla  drogas_inyectables  consumo_alcohol
##   4   2  peso  talla  drogas_inyectables  consumo_alcohol
##   4   3  peso  talla  drogas_inyectables  consumo_alcohol
##   4   4  peso  talla  drogas_inyectables  consumo_alcohol
##   4   5  peso  talla  drogas_inyectables  consumo_alcohol
##   5   1  peso  talla  drogas_inyectables  consumo_alcohol
##   5   2  peso  talla  drogas_inyectables  consumo_alcohol
##   5   3  peso  talla  drogas_inyectables  consumo_alcohol
##   5   4  peso  talla  drogas_inyectables  consumo_alcohol
##   5   5  peso  talla  drogas_inyectables  consumo_alcohol
##   6   1  peso  talla  drogas_inyectables  consumo_alcohol
##   6   2  peso  talla  drogas_inyectables  consumo_alcohol
##   6   3  peso  talla  drogas_inyectables  consumo_alcohol
##   6   4  peso  talla  drogas_inyectables  consumo_alcohol
##   6   5  peso  talla  drogas_inyectables  consumo_alcohol
##   7   1  peso  talla  drogas_inyectables  consumo_alcohol
##   7   2  peso  talla  drogas_inyectables  consumo_alcohol
##   7   3  peso  talla  drogas_inyectables  consumo_alcohol
##   7   4  peso  talla  drogas_inyectables  consumo_alcohol
##   7   5  peso  talla  drogas_inyectables  consumo_alcohol
##   8   1  peso  talla  drogas_inyectables  consumo_alcohol
##   8   2  peso  talla  drogas_inyectables  consumo_alcohol
##   8   3  peso  talla  drogas_inyectables  consumo_alcohol
##   8   4  peso  talla  drogas_inyectables  consumo_alcohol
##   8   5  peso  talla  drogas_inyectables  consumo_alcohol
##   9   1  peso  talla  drogas_inyectables  consumo_alcohol
##   9   2  peso  talla  drogas_inyectables  consumo_alcohol
##   9   3  peso  talla  drogas_inyectables  consumo_alcohol
##   9   4  peso  talla  drogas_inyectables  consumo_alcohol
##   9   5  peso  talla  drogas_inyectables  consumo_alcohol
##   10   1  peso  talla  drogas_inyectables  consumo_alcohol
##   10   2  peso  talla  drogas_inyectables  consumo_alcohol
##   10   3  peso  talla  drogas_inyectables  consumo_alcohol
##   10   4  peso  talla  drogas_inyectables  consumo_alcohol
##   10   5  peso  talla  drogas_inyectables  consumo_alcohol
##   11   1  peso  talla  drogas_inyectables  consumo_alcohol
##   11   2  peso  talla  drogas_inyectables  consumo_alcohol
##   11   3  peso  talla  drogas_inyectables  consumo_alcohol
##   11   4  peso  talla  drogas_inyectables  consumo_alcohol
##   11   5  peso  talla  drogas_inyectables  consumo_alcohol
##   12   1  peso  talla  drogas_inyectables  consumo_alcohol
##   12   2  peso  talla  drogas_inyectables  consumo_alcohol
```

```
##   12   3  peso  talla  drogas_inyectables  consumo_alcohol
##   12   4  peso  talla  drogas_inyectables  consumo_alcohol
##   12   5  peso  talla  drogas_inyectables  consumo_alcohol
##   13   1  peso  talla  drogas_inyectables  consumo_alcohol
##   13   2  peso  talla  drogas_inyectables  consumo_alcohol
##   13   3  peso  talla  drogas_inyectables  consumo_alcohol
##   13   4  peso  talla  drogas_inyectables  consumo_alcohol
##   13   5  peso  talla  drogas_inyectables  consumo_alcohol
##   14   1  peso  talla  drogas_inyectables  consumo_alcohol
##   14   2  peso  talla  drogas_inyectables  consumo_alcohol
##   14   3  peso  talla  drogas_inyectables  consumo_alcohol
##   14   4  peso  talla  drogas_inyectables  consumo_alcohol
##   14   5  peso  talla  drogas_inyectables  consumo_alcohol
##   15   1  peso  talla  drogas_inyectables  consumo_alcohol
##   15   2  peso  talla  drogas_inyectables  consumo_alcohol
##   15   3  peso  talla  drogas_inyectables  consumo_alcohol
##   15   4  peso  talla  drogas_inyectables  consumo_alcohol
##   15   5  peso  talla  drogas_inyectables  consumo_alcohol
##   16   1  peso  talla  drogas_inyectables  consumo_alcohol
##   16   2  peso  talla  drogas_inyectables  consumo_alcohol
##   16   3  peso  talla  drogas_inyectables  consumo_alcohol
##   16   4  peso  talla  drogas_inyectables  consumo_alcohol
##   16   5  peso  talla  drogas_inyectables  consumo_alcohol
##   17   1  peso  talla  drogas_inyectables  consumo_alcohol
##   17   2  peso  talla  drogas_inyectables  consumo_alcohol
##   17   3  peso  talla  drogas_inyectables  consumo_alcohol
##   17   4  peso  talla  drogas_inyectables  consumo_alcohol
##   17   5  peso  talla  drogas_inyectables  consumo_alcohol
##   18   1  peso  talla  drogas_inyectables  consumo_alcohol
##   18   2  peso  talla  drogas_inyectables  consumo_alcohol
##   18   3  peso  talla  drogas_inyectables  consumo_alcohol
##   18   4  peso  talla  drogas_inyectables  consumo_alcohol
##   18   5  peso  talla  drogas_inyectables  consumo_alcohol
##   19   1  peso  talla  drogas_inyectables  consumo_alcohol
##   19   2  peso  talla  drogas_inyectables  consumo_alcohol
##   19   3  peso  talla  drogas_inyectables  consumo_alcohol
##   19   4  peso  talla  drogas_inyectables  consumo_alcohol
##   19   5  peso  talla  drogas_inyectables  consumo_alcohol
##   20   1  peso  talla  drogas_inyectables  consumo_alcohol
##   20   2  peso  talla  drogas_inyectables  consumo_alcohol
##   20   3  peso  talla  drogas_inyectables  consumo_alcohol
##   20   4  peso  talla  drogas_inyectables  consumo_alcohol
##   20   5  peso  talla  drogas_inyectables  consumo_alcohol
##   21   1  peso  talla  drogas_inyectables  consumo_alcohol
##   21   2  peso  talla  drogas_inyectables  consumo_alcohol
##   21   3  peso  talla  drogas_inyectables  consumo_alcohol
##   21   4  peso  talla  drogas_inyectables  consumo_alcohol
##   21   5  peso  talla  drogas_inyectables  consumo_alcohol
##   22   1  peso  talla  drogas_inyectables  consumo_alcohol
##   22   2  peso  talla  drogas_inyectables  consumo_alcohol
##   22   3  peso  talla  drogas_inyectables  consumo_alcohol
##   22   4  peso  talla  drogas_inyectables  consumo_alcohol
##   22   5  peso  talla  drogas_inyectables  consumo_alcohol
##   23   1  peso  talla  drogas_inyectables  consumo_alcohol
```

```
##   23   2  peso  talla  drogas_inyectables  consumo_alcohol
##   23   3  peso  talla  drogas_inyectables  consumo_alcohol
##   23   4  peso  talla  drogas_inyectables  consumo_alcohol
##   23   5  peso  talla  drogas_inyectables  consumo_alcohol
##   24   1  peso  talla  drogas_inyectables  consumo_alcohol
##   24   2  peso  talla  drogas_inyectables  consumo_alcohol
##   24   3  peso  talla  drogas_inyectables  consumo_alcohol
##   24   4  peso  talla  drogas_inyectables  consumo_alcohol
##   24   5  peso  talla  drogas_inyectables  consumo_alcohol
##   25   1  peso  talla  drogas_inyectables  consumo_alcohol
##   25   2  peso  talla  drogas_inyectables  consumo_alcohol
##   25   3  peso  talla  drogas_inyectables  consumo_alcohol
##   25   4  peso  talla  drogas_inyectables  consumo_alcohol
##   25   5  peso  talla  drogas_inyectables  consumo_alcohol
##   26   1  peso  talla  drogas_inyectables  consumo_alcohol
##   26   2  peso  talla  drogas_inyectables  consumo_alcohol
##   26   3  peso  talla  drogas_inyectables  consumo_alcohol
##   26   4  peso  talla  drogas_inyectables  consumo_alcohol
##   26   5  peso  talla  drogas_inyectables  consumo_alcohol
##   27   1  peso  talla  drogas_inyectables  consumo_alcohol
##   27   2  peso  talla  drogas_inyectables  consumo_alcohol
##   27   3  peso  talla  drogas_inyectables  consumo_alcohol
##   27   4  peso  talla  drogas_inyectables  consumo_alcohol
##   27   5  peso  talla  drogas_inyectables  consumo_alcohol
##   28   1  peso  talla  drogas_inyectables  consumo_alcohol
##   28   2  peso  talla  drogas_inyectables  consumo_alcohol
##   28   3  peso  talla  drogas_inyectables  consumo_alcohol
##   28   4  peso  talla  drogas_inyectables  consumo_alcohol
##   28   5  peso  talla  drogas_inyectables  consumo_alcohol
##   29   1  peso  talla  drogas_inyectables  consumo_alcohol
##   29   2  peso  talla  drogas_inyectables  consumo_alcohol
##   29   3  peso  talla  drogas_inyectables  consumo_alcohol
##   29   4  peso  talla  drogas_inyectables  consumo_alcohol
##   29   5  peso  talla  drogas_inyectables  consumo_alcohol
##   30   1  peso  talla  drogas_inyectables  consumo_alcohol
##   30   2  peso  talla  drogas_inyectables  consumo_alcohol
##   30   3  peso  talla  drogas_inyectables  consumo_alcohol
##   30   4  peso  talla  drogas_inyectables  consumo_alcohol
##   30   5  peso  talla  drogas_inyectables  consumo_alcohol
##   31   1  peso  talla  drogas_inyectables  consumo_alcohol
##   31   2  peso  talla  drogas_inyectables  consumo_alcohol
##   31   3  peso  talla  drogas_inyectables  consumo_alcohol
##   31   4  peso  talla  drogas_inyectables  consumo_alcohol
##   31   5  peso  talla  drogas_inyectables  consumo_alcohol
##   32   1  peso  talla  drogas_inyectables  consumo_alcohol
##   32   2  peso  talla  drogas_inyectables  consumo_alcohol
##   32   3  peso  talla  drogas_inyectables  consumo_alcohol
##   32   4  peso  talla  drogas_inyectables  consumo_alcohol
##   32   5  peso  talla  drogas_inyectables  consumo_alcohol
##   33   1  peso  talla  drogas_inyectables  consumo_alcohol
##   33   2  peso  talla  drogas_inyectables  consumo_alcohol
##   33   3  peso  talla  drogas_inyectables  consumo_alcohol
##   33   4  peso  talla  drogas_inyectables  consumo_alcohol
##   33   5  peso  talla  drogas_inyectables  consumo_alcohol
```

```
##   34   1  peso  talla  drogas_inyectables  consumo_alcohol
##   34   2  peso  talla  drogas_inyectables  consumo_alcohol
##   34   3  peso  talla  drogas_inyectables  consumo_alcohol
##   34   4  peso  talla  drogas_inyectables  consumo_alcohol
##   34   5  peso  talla  drogas_inyectables  consumo_alcohol
##   35   1  peso  talla  drogas_inyectables  consumo_alcohol
##   35   2  peso  talla  drogas_inyectables  consumo_alcohol
##   35   3  peso  talla  drogas_inyectables  consumo_alcohol
##   35   4  peso  talla  drogas_inyectables  consumo_alcohol
##   35   5  peso  talla  drogas_inyectables  consumo_alcohol
##   36   1  peso  talla  drogas_inyectables  consumo_alcohol
##   36   2  peso  talla  drogas_inyectables  consumo_alcohol
##   36   3  peso  talla  drogas_inyectables  consumo_alcohol
##   36   4  peso  talla  drogas_inyectables  consumo_alcohol
##   36   5  peso  talla  drogas_inyectables  consumo_alcohol
##   37   1  peso  talla  drogas_inyectables  consumo_alcohol
##   37   2  peso  talla  drogas_inyectables  consumo_alcohol
##   37   3  peso  talla  drogas_inyectables  consumo_alcohol
##   37   4  peso  talla  drogas_inyectables  consumo_alcohol
##   37   5  peso  talla  drogas_inyectables  consumo_alcohol
##   38   1  peso  talla  drogas_inyectables  consumo_alcohol
##   38   2  peso  talla  drogas_inyectables  consumo_alcohol
##   38   3  peso  talla  drogas_inyectables  consumo_alcohol
##   38   4  peso  talla  drogas_inyectables  consumo_alcohol
##   38   5  peso  talla  drogas_inyectables  consumo_alcohol
##   39   1  peso  talla  drogas_inyectables  consumo_alcohol
##   39   2  peso  talla  drogas_inyectables  consumo_alcohol
##   39   3  peso  talla  drogas_inyectables  consumo_alcohol
##   39   4  peso  talla  drogas_inyectables  consumo_alcohol
##   39   5  peso  talla  drogas_inyectables  consumo_alcohol
##   40   1  peso  talla  drogas_inyectables  consumo_alcohol
##   40   2  peso  talla  drogas_inyectables  consumo_alcohol
##   40   3  peso  talla  drogas_inyectables  consumo_alcohol
##   40   4  peso  talla  drogas_inyectables  consumo_alcohol
##   40   5  peso  talla  drogas_inyectables  consumo_alcohol
##   41   1  peso  talla  drogas_inyectables  consumo_alcohol
##   41   2  peso  talla  drogas_inyectables  consumo_alcohol
##   41   3  peso  talla  drogas_inyectables  consumo_alcohol
##   41   4  peso  talla  drogas_inyectables  consumo_alcohol
##   41   5  peso  talla  drogas_inyectables  consumo_alcohol
##   42   1  peso  talla  drogas_inyectables  consumo_alcohol
##   42   2  peso  talla  drogas_inyectables  consumo_alcohol
##   42   3  peso  talla  drogas_inyectables  consumo_alcohol
##   42   4  peso  talla  drogas_inyectables  consumo_alcohol
##   42   5  peso  talla  drogas_inyectables  consumo_alcohol
##   43   1  peso  talla  drogas_inyectables  consumo_alcohol
##   43   2  peso  talla  drogas_inyectables  consumo_alcohol
##   43   3  peso  talla  drogas_inyectables  consumo_alcohol
##   43   4  peso  talla  drogas_inyectables  consumo_alcohol
##   43   5  peso  talla  drogas_inyectables  consumo_alcohol
##   44   1  peso  talla  drogas_inyectables  consumo_alcohol
##   44   2  peso  talla  drogas_inyectables  consumo_alcohol
##   44   3  peso  talla  drogas_inyectables  consumo_alcohol
##   44   4  peso  talla  drogas_inyectables  consumo_alcohol
```

```
##   44   5  peso  talla  drogas_inyectables  consumo_alcohol
##   45   1  peso  talla  drogas_inyectables  consumo_alcohol
##   45   2  peso  talla  drogas_inyectables  consumo_alcohol
##   45   3  peso  talla  drogas_inyectables  consumo_alcohol
##   45   4  peso  talla  drogas_inyectables  consumo_alcohol
##   45   5  peso  talla  drogas_inyectables  consumo_alcohol
##   46   1  peso  talla  drogas_inyectables  consumo_alcohol
##   46   2  peso  talla  drogas_inyectables  consumo_alcohol
##   46   3  peso  talla  drogas_inyectables  consumo_alcohol
##   46   4  peso  talla  drogas_inyectables  consumo_alcohol
##   46   5  peso  talla  drogas_inyectables  consumo_alcohol
##   47   1  peso  talla  drogas_inyectables  consumo_alcohol
##   47   2  peso  talla  drogas_inyectables  consumo_alcohol
##   47   3  peso  talla  drogas_inyectables  consumo_alcohol
##   47   4  peso  talla  drogas_inyectables  consumo_alcohol
##   47   5  peso  talla  drogas_inyectables  consumo_alcohol
##   48   1  peso  talla  drogas_inyectables  consumo_alcohol
##   48   2  peso  talla  drogas_inyectables  consumo_alcohol
##   48   3  peso  talla  drogas_inyectables  consumo_alcohol
##   48   4  peso  talla  drogas_inyectables  consumo_alcohol
##   48   5  peso  talla  drogas_inyectables  consumo_alcohol
##   49   1  peso  talla  drogas_inyectables  consumo_alcohol
##   49   2  peso  talla  drogas_inyectables  consumo_alcohol
##   49   3  peso  talla  drogas_inyectables  consumo_alcohol
##   49   4  peso  talla  drogas_inyectables  consumo_alcohol
##   49   5  peso  talla  drogas_inyectables  consumo_alcohol
##   50   1  peso  talla  drogas_inyectables  consumo_alcohol
##   50   2  peso  talla  drogas_inyectables  consumo_alcohol
##   50   3  peso  talla  drogas_inyectables  consumo_alcohol
##   50   4  peso  talla  drogas_inyectables  consumo_alcohol
##   50   5  peso  talla  drogas_inyectables  consumo_alcohol
```

```r
datos_imputados <- complete(imp)

# Convertir de nuevo a factores las variables categóricas
datos_modelo_completo <- datos_imputados %>%
    mutate(
        depresion = factor(depresion, levels=c(1,2), labels=c("Si", "No")),
        sexo_al_nacer = factor(sexo_al_nacer, levels=c(1,2), labels=c("Hombre", "Mujer")),
        nivel_educacional = factor(nivel_educacional),
        bienestar_emocional = factor(bienestar_emocional),
        calidad_vida_percibida = factor(calidad_vida_percibida),
        satisfaccion_aspecto_fisico = factor(satisfaccion_aspecto_fisico),
        drogas_inyectables = factor(drogas_inyectables, levels=c(1,2), labels=c("Si", "No")),
        consumo_alcohol = factor(consumo_alcohol)
    )

# Dividir en conjunto de entrenamiento y prueba
set.seed(123)
index_train <- createDataPartition(datos_modelo_completo$depresion, p = 0.7, list = FALSE)
train_data <- datos_modelo_completo[index_train,]
test_data <- datos_modelo_completo[-index_train,]

# Verificar distribución en conjuntos de entrenamiento y prueba
```

```r
print("Distribución en conjunto de entrenamiento:")
```

```
## [1] "Distribución en conjunto de entrenamiento:"
```

```r
print(table(train_data$depresion))
```

```
##
##   Si   No
##  536 1387
```

```r
print("Distribución en conjunto de prueba:")
```

```
## [1] "Distribución en conjunto de prueba:"
```

```r
print(table(test_data$depresion))
```

```
##
##  Si  No
## 229 594
```

```r
# Configurar control de entrenamiento con validación cruzada estratificada
ctrl <- trainControl(
    method = "repeatedcv",
    number = 5,
    repeats = 3,
    classProbs = TRUE,
    summaryFunction = twoClassSummary,
    sampling = "down"  # Usar doum-sampling en lugar de SMOTE
)

# Modelo 1: Random Forest con hiperparámetros optimizados
set.seed(123)
modelo_rf <- train(
    depresion ~ .,
    data = train_data,
    method = "rf",
    trControl = ctrl,
    metric = "ROC",
    tuneLength = 10
)

# Modelo 2: Árbol de Decisión con hiperparámetros optimizados
set.seed(123)
modelo_tree <- train(
    depresion ~ .,
    data = train_data,
    method = "rpart",
    trControl = ctrl,
    metric = "ROC",
    tuneLength = 10
```

```
)

# Evaluar modelos en conjunto de prueba
pred_rf <- predict(modelo_rf, test_data)
pred_tree <- predict(modelo_tree, test_data)

# Métricas para Random Forest
cm_rf <- confusionMatrix(pred_rf, test_data$depresion)
print("Métricas para Random Forest en datos de prueba:")
```

```
## [1] "Métricas para Random Forest en datos de prueba:"
```

```
print(cm_rf)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  Si  No
##         Si 156 337
##         No  73 257
##
##                Accuracy : 0.5018
##                  95% CI : (0.4671, 0.5365)
##     No Information Rate : 0.7217
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0841
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.6812
##             Specificity : 0.4327
##          Pos Pred Value : 0.3164
##          Neg Pred Value : 0.7788
##              Prevalence : 0.2783
##          Detection Rate : 0.1896
##    Detection Prevalence : 0.5990
##       Balanced Accuracy : 0.5569
##
##        'Positive' Class : Si
##
```

```
# Métricas para Árbol de Decisión
cm_tree <- confusionMatrix(pred_tree, test_data$depresion)
print("Métricas para Árbol de Decisión en datos de prueba:")
```

```
## [1] "Métricas para Árbol de Decisión en datos de prueba:"
```

```
print(cm_tree)
```

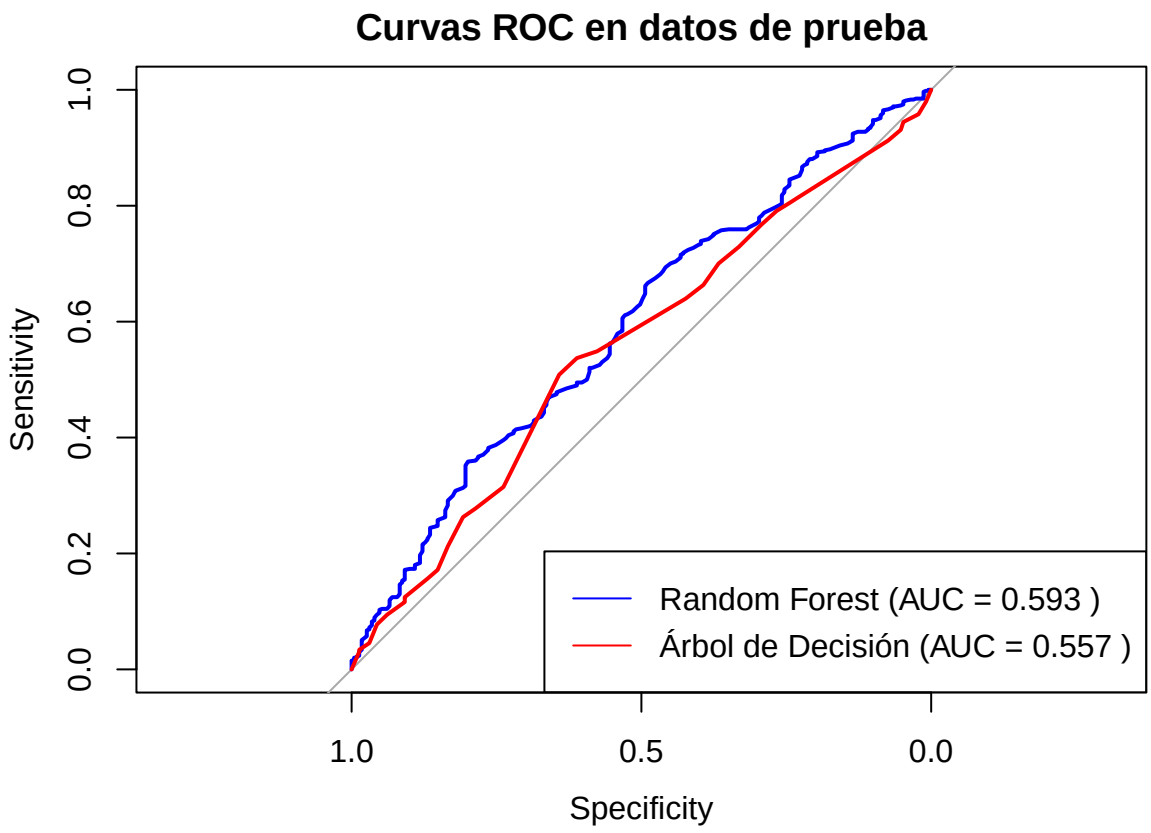
```
## Confusion Matrix and Statistics
```

```
##
##           Reference
## Prediction  Si  No
##         Si 132 268
##         No  97 326
##
##                Accuracy : 0.5565
##                  95% CI : (0.5218, 0.5908)
##     No Information Rate : 0.7217
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.1019
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.5764
##             Specificity : 0.5488
##          Pos Pred Value : 0.3300
##          Neg Pred Value : 0.7707
##              Prevalence : 0.2783
##          Detection Rate : 0.1604
##    Detection Prevalence : 0.4860
##       Balanced Accuracy : 0.5626
##
##        'Positive' Class : Si
##
```

```r
# Curvas ROC en datos de prueba
pred_rf_prob <- predict(modelo_rf, test_data, type = "prob")
roc_rf <- roc(test_data$depresion, pred_rf_prob[,"Si"])
auc_rf <- auc(roc_rf)

pred_tree_prob <- predict(modelo_tree, test_data, type = "prob")
roc_tree <- roc(test_data$depresion, pred_tree_prob[,"Si"])
auc_tree <- auc(roc_tree)

# Graficar curvas ROC
plot(roc_rf, main = "Curvas ROC en datos de prueba", col = "blue")
lines(roc_tree, col = "red")
legend("bottomright",
       legend = c(paste("Random Forest (AUC =", round(auc_rf, 3), ")"),
                  paste("Árbol de Decisión (AUC =", round(auc_tree, 3), ")")),
       col = c("blue", "red"),
       lty = 1)
```




```r
# Importancia de variables para Random Forest
print("Importancia de variables en Random Forest:")
```

```
## [1] "Importancia de variables en Random Forest:"
```

```r
varImp(modelo_rf)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 40)
##
##                               Overall
## edad                           100.00
## peso                            85.12
## talla                           78.56
## bienestar_emocional7            33.50
## bienestar_emocional4            28.70
## sexo_al_nacerMujer              28.35
## consumo_alcohol3                25.51
## bienestar_emocional3            24.41
## consumo_alcohol2                24.24
## calidad_vida_percibida5         21.55
## calidad_vida_percibida3         18.91
## satisfaccion_aspecto_fisico2    18.61
```

```
## nivel_educacional5             18.40
## satisfaccion_aspecto_fisico5   18.11
## nivel_educacional14            16.31
## nivel_educacional3             15.95
## bienestar_emocional6           12.88
## bienestar_emocional2           12.57
## bienestar_emocional5           11.99
## satisfaccion_aspecto_fisico3   11.91
```

```
# Comparación final de modelos
resultados <- data.frame(
    Modelo = c("Random Forest", "Árbol de Decisión"),
    Sensibilidad = c(cm_rf$byClass["Sensitivity"], cm_tree$byClass["Sensitivity"]),
    Especificidad = c(cm_rf$byClass["Specificity"], cm_tree$byClass["Specificity"]),
    VPP = c(cm_rf$byClass["Pos Pred Value"], cm_tree$byClass["Pos Pred Value"]),
    Exactitud = c(cm_rf$overall["Accuracy"], cm_tree$overall["Accuracy"]),
    AUC = c(auc_rf, auc_tree)
)

print("\nComparación final de modelos en datos de prueba:")
```

```
## [1] "\nComparación final de modelos en datos de prueba:"
```

```
print(resultados)
```

```
##              Modelo Sensibilidad Especificidad     VPP Exactitud       AUC
## 1     Random Forest    0.6812227     0.4326599 0.31643 0.5018226 0.5930668
## 2 Árbol de Decisión    0.5764192     0.5488215 0.33000 0.5565006 0.5565370
```

6. Conclusiones

```
output2 <- c(

  "\nResultados de los Modelos Predictivos:\n",
  "Análisis por métrica\n\n",

  "Sensibilidad (Recall):\n\n",
  "Random Forest: 68.12%. Este modelo detecta correctamente la mayoría de los casos positivos (es decir
  "Árbol de Decisión: 57.64%. Tiene un peor desempeño en la detección de casos positivos comparado con l
  "Conclusión: Random Forest es superior en identificar casos positivos.\n\n",

  "Especificidad:\n\n",
  "Random Forest: 43.27%. Este modelo identifica correctamente a los casos negativos con una precisión l
  "Árbol de Decisión: 54.88%. Muestra mejor capacidad para identificar correctamente a los casos negativ
  "Conclusión: El Árbol de Decisión supera a Random Forest en esta métrica.\n\n",

  "Valor Predictivo Positivo (VPP):\n\n",
  "Random Forest: 31.64%. De todos los casos que predice como positivos, solo el 31.64% son verdaderos p
  "Árbol de Decisión: 33.00%. Tiene un desempeño ligeramente mejor en este aspecto.\n\n",
  "Conclusión: Ambos modelos tienen limitaciones en la predicción positiva, aunque el Árbol de Decisión
```

```
  "Exactitud:\n\n",
  "Random Forest: 50.18%. Este modelo tiene un desempeño cercano al azar en la clasificación general.\n\
  "Árbol de Decisión: 55.65%. Supera ligeramente a Random Forest en esta métrica.\n\n",
  "Conclusión: Ninguno de los modelos tiene un desempeño excelente, pero el Árbol de Decisión es margina

  "AUC (Área bajo la curva ROC):\n\n",
  "Random Forest: 59.31%. Tiene un desempeño aceptable pero limitado en términos de discriminación entre
  "Árbol de Decisión: 55.65%. Tiene un desempeño inferior a Random Forest, con una discriminación cercan
  "Conclusión: Random Forest tiene una mejor capacidad general para distinguir entre positivos y negativ

  "Consideraciones y Limitaciones:\n",
  "  - Desbalance de Clases: mayor proporción de 'No' que de 'Sí', afectando el desempeño de los modelos
  "  - Calidad de datos: datos faltantes impactan la representatividad, especialmente en variables como
  "  - Selección de Variables: se recomienda análisis de correlación para mejorar la calidad del modelo
  "  - Validación Cruzada: es crucial que el conjunto de prueba sea representativo y mantenga la distrib

  "Mejoras Futuras:\n",
  "  - Exploración de Otros Modelos: probar otros algoritmos de aprendizaje automático (SVM, XGBoost, et
  "  - Análisis de Importancia de Variables: investigar qué variables influyen en la predicción de la de
  "  - Recoger Más Datos: recolectar más datos para mejorar la robustez y capacidad de generalización de

  "Conclusiones generales:\n",
  "  - Los modelos predictivos desarrollados a partir de la encuesta ENSSEX 2022-2023 brindan informacio
  "  - Se requiere un mayor esfuerzo para mejorar la precisión de la identificación de casos reales.\n"
  "  - La colaboración entre expertos en epidemiología y data science puede facilitar el desarrollo de m
)

cat(output2, sep = "")
```

```
##
## Resultados de los Modelos Predictivos:
## Análisis por métrica
##
## Sensibilidad (Recall):
##
## Random Forest: 68.12%. Este modelo detecta correctamente la mayoría de los casos positivos (es decir
##
## Árbol de Decisión: 57.64%. Tiene un peor desempeño en la detección de casos positivos comparado con l
##
## Conclusión: Random Forest es superior en identificar casos positivos.
##
## Especificidad:
##
## Random Forest: 43.27%. Este modelo identifica correctamente a los casos negativos con una precisión l
##
## Árbol de Decisión: 54.88%. Muestra mejor capacidad para identificar correctamente a los casos negativ
##
## Conclusión: El Árbol de Decisión supera a Random Forest en esta métrica.
##
## Valor Predictivo Positivo (VPP):
##
## Random Forest: 31.64%. De todos los casos que predice como positivos, solo el 31.64% son verdaderos p
##
```

```
## Árbol de Decisión: 33.00%. Tiene un desempeño ligeramente mejor en este aspecto.
##
## Conclusión: Ambos modelos tienen limitaciones en la predicción positiva, aunque el Árbol de Decisión
##
## Exactitud:
##
## Random Forest: 50.18%. Este modelo tiene un desempeño cercano al azar en la clasificación general.
##
## Árbol de Decisión: 55.65%. Supera ligeramente a Random Forest en esta métrica.
##
## Conclusión: Ninguno de los modelos tiene un desempeño excelente, pero el Árbol de Decisión es margina
##
## AUC (Área bajo la curva ROC):
##
## Random Forest: 59.31%. Tiene un desempeño aceptable pero limitado en términos de discriminación entre
##
## Árbol de Decisión: 55.65%. Tiene un desempeño inferior a Random Forest, con una discriminación cercan
##
## Conclusión: Random Forest tiene una mejor capacidad general para distinguir entre positivos y negativ
##
## Consideraciones y Limitaciones:
##   - Desbalance de Clases: mayor proporción de 'No' que de 'Sí', afectando el desempeño de los modelos
##   - Calidad de datos: datos faltantes impactan la representatividad, especialmente en variables como
##   - Selección de Variables: se recomienda análisis de correlación para mejorar la calidad del modelo
##   - Validación Cruzada: es crucial que el conjunto de prueba sea representativo y mantenga la distrib
##
## Mejoras Futuras:
##   - Exploración de Otros Modelos: probar otros algoritmos de aprendizaje automático (SVM, XGBoost, et
##   - Análisis de Importancia de Variables: investigar qué variables influyen en la predicción de la de
##   - Recoger Más Datos: recolectar más datos para mejorar la robustez y capacidad de generalización de
##
## Conclusiones generales:
##   - Los modelos predictivos desarrollados a partir de la encuesta ENSSEX 2022-2023 brindan informaci
##   - Se requiere un mayor esfuerzo para mejorar la precisión de la identificación de casos reales.
##   - La colaboración entre expertos en epidemiología y data science puede facilitar el desarrollo de m
```