

INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
**PRACTICA DE CLASIFICACIÓN (Bagging,
Boosting)**

Grupo: 5AM1

No. Equipo: 2

Nombres:

Ulises Abdiel Cabello Cardenas

Fernanda Madariaga Villanueva

Alejandro Maldonado López

Andrés Zamarripa Altamirano



Figure 1: Enter Caption

Objetivo de la práctica

Objetivo: Crear un modelo de clasificación para identificar las características distintivas en los tipos de voz de mujeres y de hombres. Aplique métodos de Random Forest (Bagging) y Boosting.

Descripción del conjunto de datos

Reconocimiento de Género por Voz. Identificar una voz como masculina o femenina.

No. Licencia: CC BY-NC-SA 4.0

Disponible en: <https://www.kaggle.com/datasets/primaryobjects/voicegender/data>

Diccionario de datos

El conjunto de datos cuenta con 22 atributos, los cuales se describen en la siguiente tabla.

Variable	Tipo	Significado
meanfreq	numérico	Frecuencia media (kHz)
sd	numérico	Desviación estándar de la frecuencia
median	numérico	Frecuencia mediana (kHz)
Q25	numérico	Primer cuantil (kHz)
Q75	numérico	Tercer cuantil (kHz)
IQR	numérico	Rango intercuartílico (kHz)
skew	numérico	Asimetría del espectro
kurt	numérico	Curtosis del espectro
sp.ent	numérico	Entropía espectral
sfm	numérico	Planitud espectral
mode	numérico	Frecuencia de la moda
centroid	numérico	Centroide de frecuencia
peakf	numérico	Frecuencia pico (mayor energía)
meanfun	numérico	Promedio de frecuencia fundamental
minfun	numérico	Frecuencia fundamental mínima
maxfun	numérico	Frecuencia fundamental máxima
meandom	numérico	Promedio de frecuencia dominante
mindom	numérico	Frecuencia dominante mínima
maxdom	numérico	Frecuencia dominante máxima
dfrange	numérico	Rango de frecuencia dominante
modindx	numérico	Índice de modulación
label	categorico	Género: masculino o femenino

A continuacion se describiran las variables con mayor detalle explicando su significado.

meanfreq Frecuencia media (kHz). Es el promedio de todas las frecuencias presentes en el espectro de la señal de voz. Representa la tendencia central del contenido espectral. En términos de género, voces femeninas suelen presentar frecuencias medias más

altas que las voces masculinas debido a diferencias fisiológicas en el aparato fonador `rabiner'speech`.

sd Desviación estándar de la frecuencia. Mide la dispersión de las frecuencias alrededor de la frecuencia media. Valores grandes indican que la energía de la voz está distribuida en un rango amplio de frecuencias, mientras que valores pequeños sugieren un espectro más concentrado.

median Frecuencia mediana (kHz). Es la frecuencia que divide la distribución espectral en dos partes con igual energía acumulada. A diferencia de la media, la mediana es menos sensible a valores extremos y describe de forma robusta el “centro” del espectro.

Q25 Primer cuartil (kHz). Frecuencia por debajo de la cual se encuentra el 25% de la energía del espectro. Da información sobre la parte baja del rango de frecuencias donde se concentra la energía de la voz.

Q75 Tercer cuartil (kHz). Frecuencia por debajo de la cual se encuentra el 75% de la energía del espectro. Junto con Q25 permite describir la dispersión del contenido espectral.

IQR Rango intercuartílico (kHz). Definido como $IQR = Q75 - Q25$. Mide la amplitud del rango central de frecuencias donde se concentra la mayor parte de la energía, ignorando los extremos. Un IQR grande indica un espectro más extendido.

skew Asimetría del espectro. Indica si la distribución de energía en frecuencia está sesgada hacia valores altos o bajos: valores positivos implican mayor cola hacia frecuencias altas, mientras que valores negativos implican una cola hacia frecuencias bajas. La asimetría está relacionada con la forma del timbre de la voz `seewave'pkg`.

kurt Curtosis del espectro. Describe qué tan “picuda” o “aplanada” es la distribución espectral en comparación con una distribución normal. Valores altos de curtosis indican que la energía se concentra en pocas frecuencias (picos espectrales pronunciados), en tanto que valores bajos corresponden a un espectro más uniforme.

sp.ent Entropía espectral. Mide el grado de desorden o aleatoriedad del espectro de magnitud. Valores altos indican un espectro más uniforme (similar a ruido), mientras que valores bajos corresponden a espectros más estructurados, típicos de señales periódicas como la voz `tchernichovski'entropy`.

sfm Planitud espectral (Spectral Flatness Measure). Relación entre la media geométrica y la media aritmética del espectro. Cuando la planitud se acerca a 1, la señal se parece a ruido blanco (energía distribuida de forma casi uniforme). Valores cercanos a 0 indican que el espectro está dominado por unos pocos picos (tono más definido).

mode Frecuencia de la moda. Es la frecuencia con mayor energía en el espectro (el pico dominante). Puede asociarse a una frecuencia resonante importante de la voz y complementa la información de `meanfreq` y `centroid`.

centroid Centroide espectral. Corresponde al “centro de masa” del espectro, calculado como un promedio ponderado de las frecuencias por su energía. Suele relacionarse con la percepción de brillo del sonido: centroids más altos se perciben como voces más agudas **tzanetakis’speech**.

peakf Frecuencia pico. Frecuencia asociada al máximo valor de energía en el espectro. En muchos casos coincide con una de las formantes o con la frecuencia dominante de la señal.

meanfun Frecuencia fundamental promedio. Promedio de la frecuencia fundamental F_0 a lo largo de la grabación. Es una de las variables más importantes para diferenciar género, pues típicamente las voces femeninas poseen frecuencias fundamentales más altas que las masculinas **rabiner’speech**.

minfun Frecuencia fundamental mínima. Valor mínimo de F_0 observado en la señal. Refleja el tono más grave producido durante la grabación.

maxfun Frecuencia fundamental máxima. Valor máximo de F_0 observado en la señal, relacionado con el tono más agudo alcanzado por la persona hablante.

meandom Frecuencia dominante promedio. Promedio de la frecuencia dominante (la de mayor energía) calculada en cada ventana temporal. Aunque se relaciona con la frecuencia fundamental, puede capturar también formantes y otras componentes relevantes del habla.

mindom Frecuencia dominante mínima. Menor valor de frecuencia dominante registrado en las diferentes ventanas temporales del análisis.

maxdom Frecuencia dominante máxima. Mayor frecuencia dominante observada en la señal. Junto con **mindom** resume el rango de variación de la frecuencia dominante.

dfrange Rango de frecuencia dominante. Diferencia entre la frecuencia dominante máxima y mínima. Cuantifica la variación global de la frecuencia dominante a lo largo de la grabación, reflejando cambios en el tono y en la energía del habla.

modindx Índice de modulación. Se calcula como la suma acumulada de las diferencias absolutas entre mediciones consecutivas de la frecuencia fundamental, dividida entre el rango total de dicha frecuencia. Un **modindx** alto indica una voz con variaciones frecuentes de tono (entonación más dinámica), mientras que valores bajos corresponden a voces más monótonas.

label Género (variable objetivo). Variable categórica que indica el género de la persona hablante: **male** o **female**. Es la variable dependiente del problema de clasificación; todas las demás variables actúan como predictores para construir el modelo.

Consideraciones encontradas en el conjunto de datos

Describir las consideraciones que encuentre en el conjunto de datos.

Name	Type	# Missing values	# Unique values
sd	Number (Float)	0	3166
median	Number (Float)	0	3077
Q25	Number (Float)	0	3103
Q75	Number (Float)	0	3034
IQR	Number (Float)	0	3073
skew	Number (Float)	0	3166
kurt	Number (Float)	0	3166
sp.ent	Number (Float)	0	3166
sfm	Number (Float)	0	3166
mode	Number (Float)	0	2825
centroid	Number (Float)	0	3166

Figure 2: Consideraciones del conjunto de datos

meanfun	Number (Float)	0	3166
minfun	Number (Float)	0	913
maxfun	Number (Float)	0	123
meandom	Number (Float)	0	2999
mindom	Number (Float)	0	77
maxdom	Number (Float)	0	1054
dfrange	Number (Float)	0	1091
modindx	Number (Float)	0	3079
meanfreq	Number (Float)	0	3166

Figure 3: Consideraciones del conjunto de datos

En primera instancia nos dimos cuenta de que hace falta la dimensión de peakf (Frecuencia pico – Mayor energía), el dataset tiene una dimensión de 3168 filas por 21 columnas de las cuales 20 son numéricas y una es categórica la cual es nuestra variable objetivo (male y female). Esto facilita el uso de modelos basados en árboles, pues no requieren codificación adicional ni normalización estricta.

- **Tipos de datos:** Todos los atributos predictivos del conjunto de datos son de tipo *numérico (float)*, mientras que la variable objetivo `label` es *categórica* con dos clases: `male` y `female`. Esto facilita el uso de modelos basados en árboles, pues no requieren codificación adicional ni normalización estricta.
- **Valores faltantes:** Ninguna variable contiene valores faltantes: Missing Values = 0. Por lo tanto, no se necesitó realizar imputación ni eliminación de registros.
- **Número de valores únicos:** Varias variables presentan un número muy alto de valores únicos (hasta 3166), lo que indica un comportamiento continuo. Sin embargo, se observaron variables con menor variabilidad, tales como:
 - `maxfun`: 123 valores únicos
 - `mindom`: 77 valores únicos
 - `minfun`: 913 valores únicos
 - `dfrange`: 1091 valores únicos

Estas diferencias sugieren que algunas características espectrales poseen menor dispersión que otras.

- **Presencia de outliers:** Aunque no existen valores faltantes, se identifica la presencia de valores extremos en variables como: `skew`, `kurt`, `maxdom` y `dfrange`. Debido a que los modelos de tipo árbol (Random Forest y Boosting) son robustos ante outliers, se decidió no realizar eliminación ni transformación de estos valores.
- **Balance de clases:** El conjunto de datos está perfectamente balanceado con: 1584 voces masculinas y 1584 voces femeninas. Esto evita la necesidad de aplicar técnicas de balanceo como *oversampling* o *undersampling*.
- **Distribución y variabilidad de los atributos:** Atributos como `meanfreq`, `median`, `meanfun` y `centroid` presentan alta variabilidad, lo que corresponde con mediciones continuas del espectro de voz. En contraste, variables como `mindom` y `maxdom` muestran menos valores únicos debido a la naturaleza discretizada de la frecuencia dominante.

Matriz de correlacion

Para comprender las relaciones lineales entre las variables predictoras de nuestro dataset, analizamos la matriz de correlación. Este análisis es crucial, ya que un alto grado de correlación entre características, conocido como multicolinealidad, puede afectar el rendimiento y la interpretabilidad de nuestros modelos de Bagging y Boosting, la matriz de correlación es la siguiente:

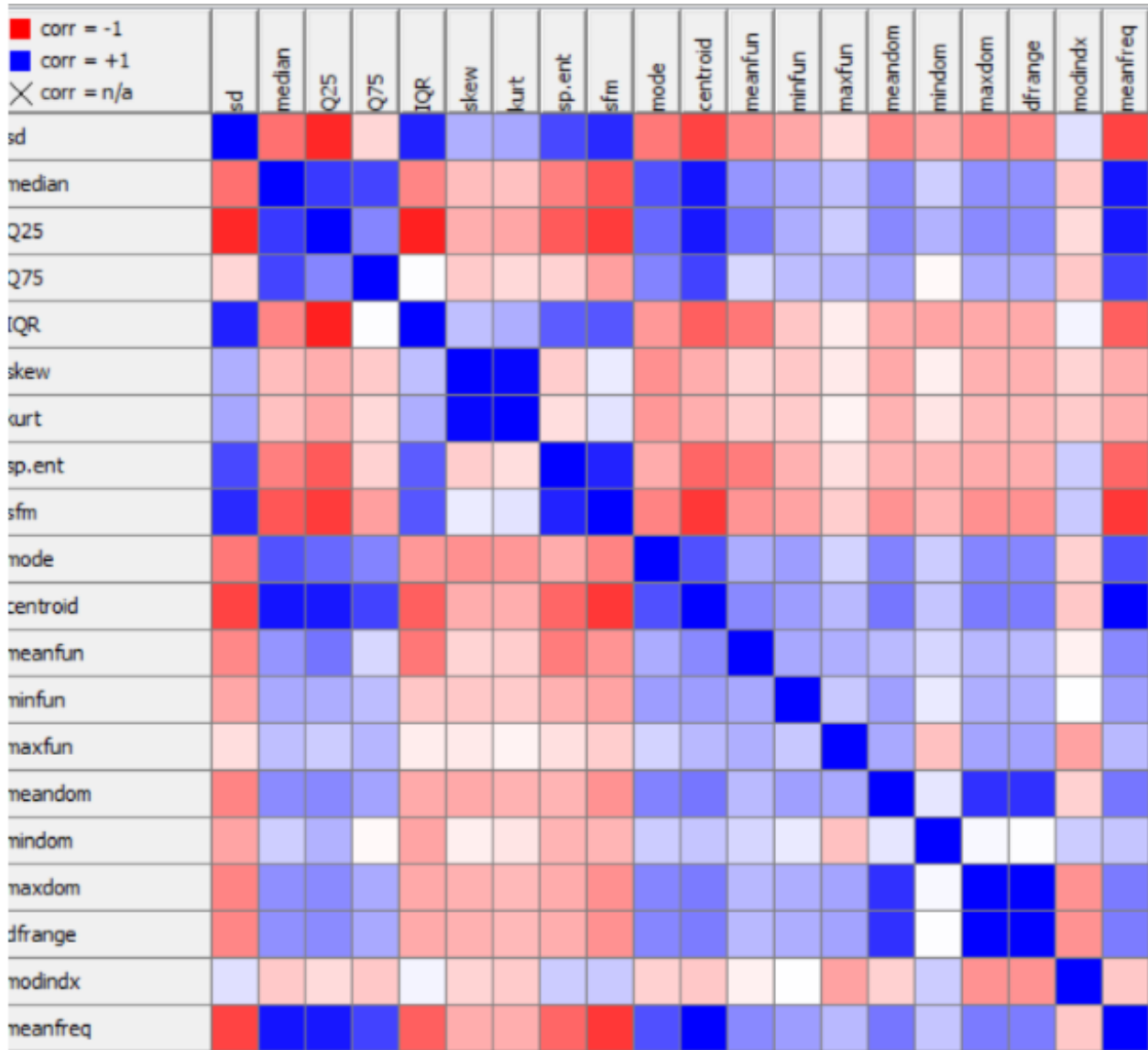


Figure 4: Correlacion matrix

Al observar la matriz, identificamos que la mayoría de las variables presentan correlaciones bajas o moderadas, lo que es beneficioso para el modelo, ya que sugiere que cada característica aporta información única. Sin embargo, se detectan algunos pares de variables con una correlación positiva muy alta, cercana a +1. Esto era esperable, ya que varias de estas características describen propiedades estadísticas intrínsecamente relacionadas de una misma distribución de frecuencias. Por ejemplo, es natural encontrar una fuerte correlación entre los cuartiles (Q25 y Q75) y la mediana, o entre el rango intercuartílico (IQR) y dichos cuartiles, ya que una se calcula directamente a partir de las otras.

Estos hallazgos nos indican que, si bien el conjunto de datos es rico en información, existe una redundancia inherente en algunas de sus dimensiones. Si bien algoritmos como Random Forest (Bagging) y Gradient Boosting pueden ser robustos a cierto nivel de correlación, este

conocimiento es valioso para etapas posteriores. En caso de buscar simplificar el modelo, podríamos considerar la posibilidad de eliminar algunas variables altamente correlacionadas o aplicar técnicas de reducción de dimensionalidad, como el Análisis de Componentes Principales (PCA), para consolidar esta información superpuesta en componentes independientes sin perder la esencia de la información acústica original.

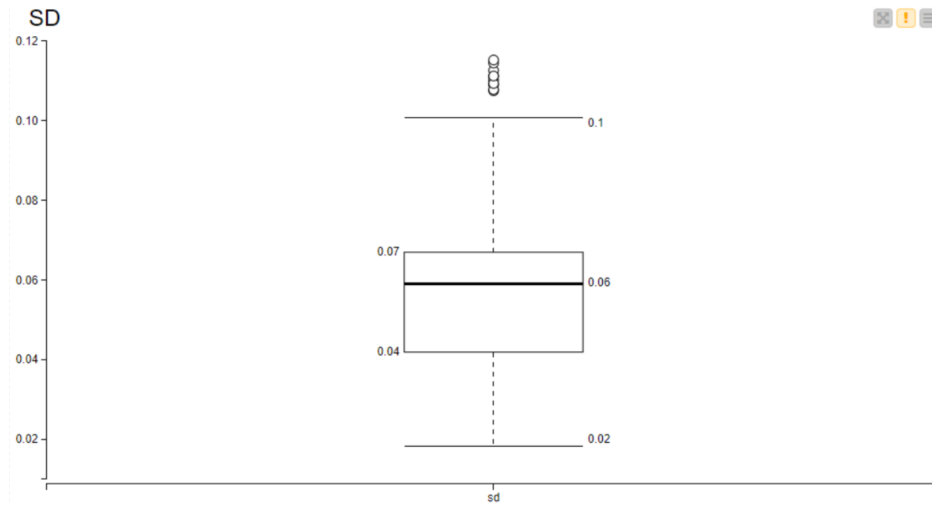


Figure 5: Caja de la variable SD(standar deviation

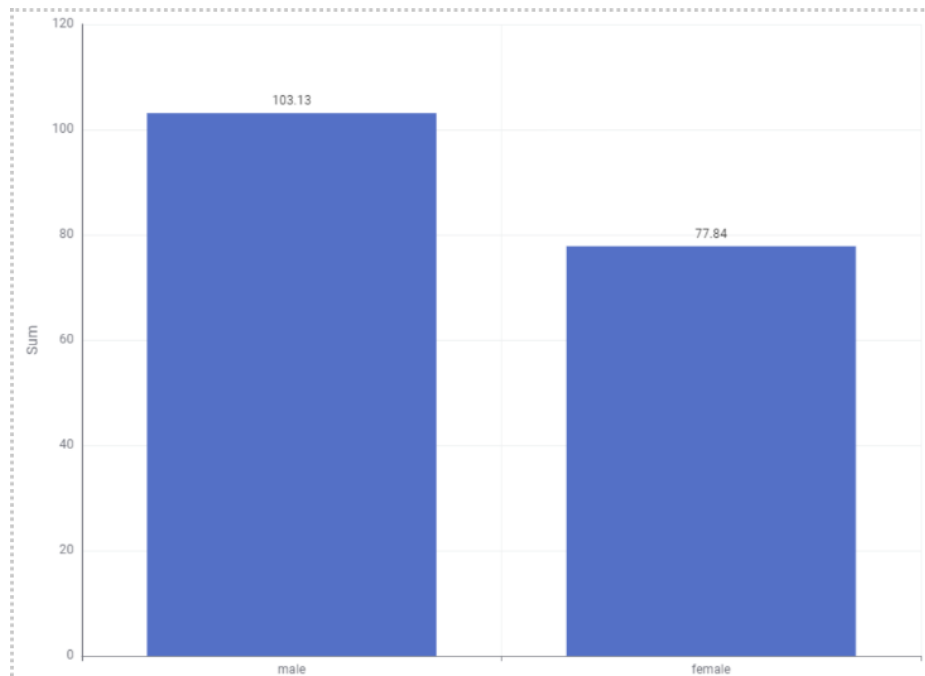


Figure 6: Grafico de barras de sd por género

Tras el proceso de normalización de los datos, procedimos a realizar un análisis de distribución mediante diagramas de caja (boxplots) e histogramas agrupados por género para

cada variable. El análisis de la variable *sd*, que representa la desviación estándar de la frecuencia, revela patrones distintivos entre las clases.

El diagrama de caja muestra que la distribución de la desviación estándar de la frecuencia es notablemente diferente para "male" y "female". Se observa que las muestras etiquetadas como "male" presentan una mediana más alta y una caja en una posición general superior en la escala, lo que indica que los valores de esta variable son consistentemente mayores para el género masculino. Además, la presencia de numerosos puntos atípicos (outliers) por encima del bigote superior en ambas categorías sugiere una considerable variabilidad en los datos, siendo esta dispersión más pronunciada en el caso de las voces masculinas.

Esta observación se corrobora y cuantifica en el gráfico de barras. En él, se aprecia con claridad que la media de la variable "sd" para "male" es significativamente mayor (103.13) en comparación con la media para "female" (77.84). Está marcada diferencia de aproximadamente 25 unidades en la desviación estándar de la frecuencia entre géneros posiciona a esta variable como un predictor potencialmente muy fuerte para nuestro modelo de clasificación. La clara separación en las medidas de tendencia central sugiere que la dispersión espectral es una característica acústica fundamental para discriminar entre voces masculinas y femeninas.

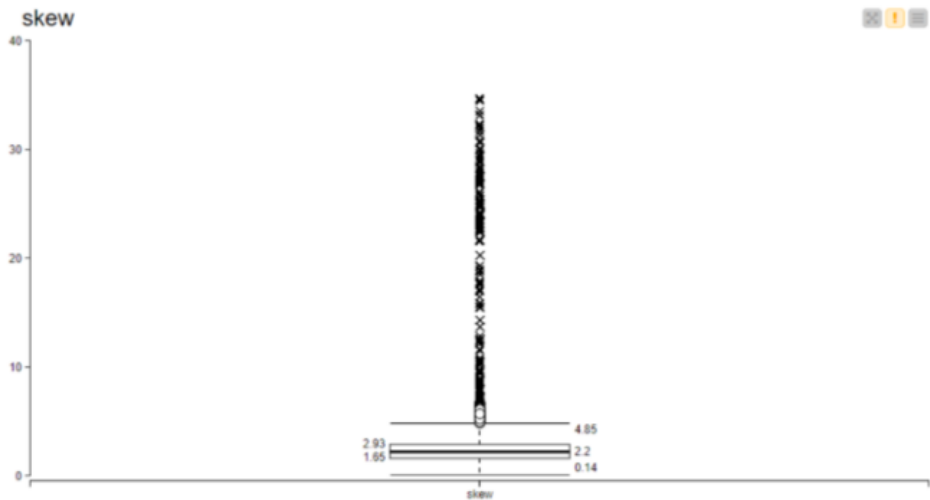


Figura 6. Caja de la variable "skew".

Figure 7: Caja de la variable "skew"

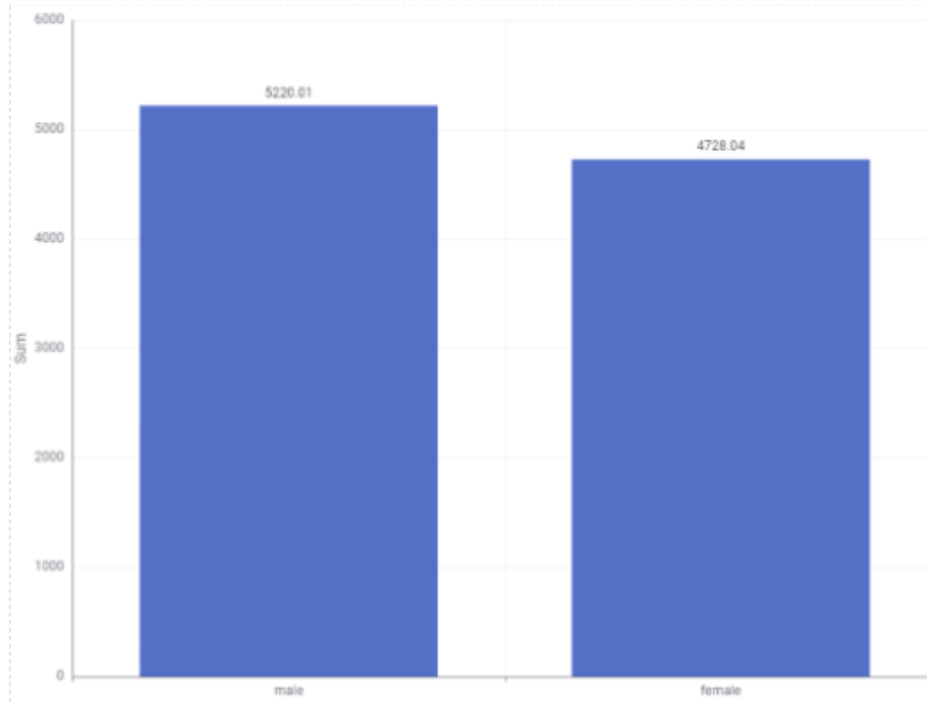


Figure 8: Gráfico de barras de “skew” por género.

El análisis de la asimetría espectral (“skew”) revela una distribución de datos extremadamente sesgada, como lo evidencia el histograma. La gran mayoría de las observaciones se concentran en un rango muy estrecho de valores bajos, cercano a cero, creando una barra dominante. Sin embargo, la presencia de una cola larga hacia la derecha, con valores de “skew” que se extienden hasta 40 y más allá, es la característica más destacable.

Esta distribución no normal se confirma en el diagrama de caja, donde la caja principal aparece comprimida en la parte inferior del gráfico, mientras que los bigotes se extienden de manera excepcionalmente larga, particularmente hacia los valores positivos. La mediana para “male” se sitúa en 5220, un valor inusualmente alto que parece estar fuertemente influenciado por la presencia de valores extremos en esta categoría. Este patrón indica que “skew” es una variable con una alta varianza y una distribución muy asimétrica, lo que podría requerir una transformación adicional o ser manejada con cuidado por los modelos para evitar un impacto desproporcionado de los outliers en la clasificación.

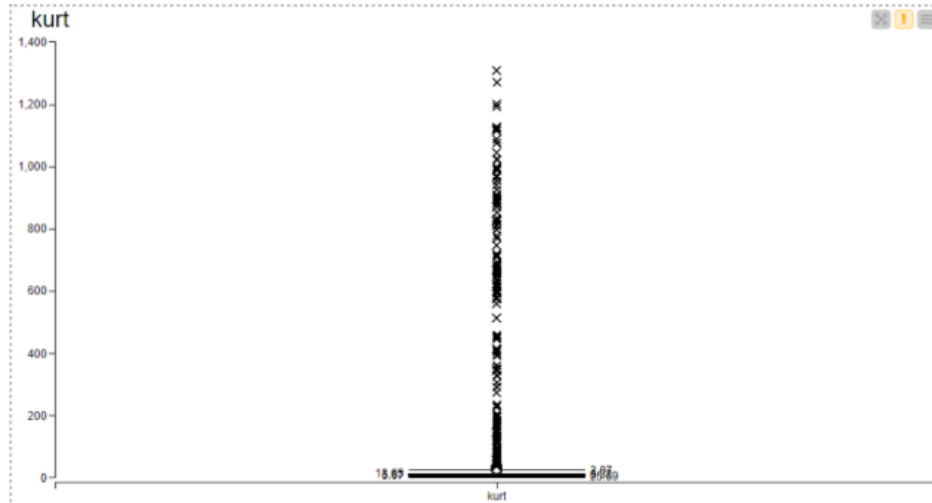


Figure 9: Caja de la variable “kurt”.

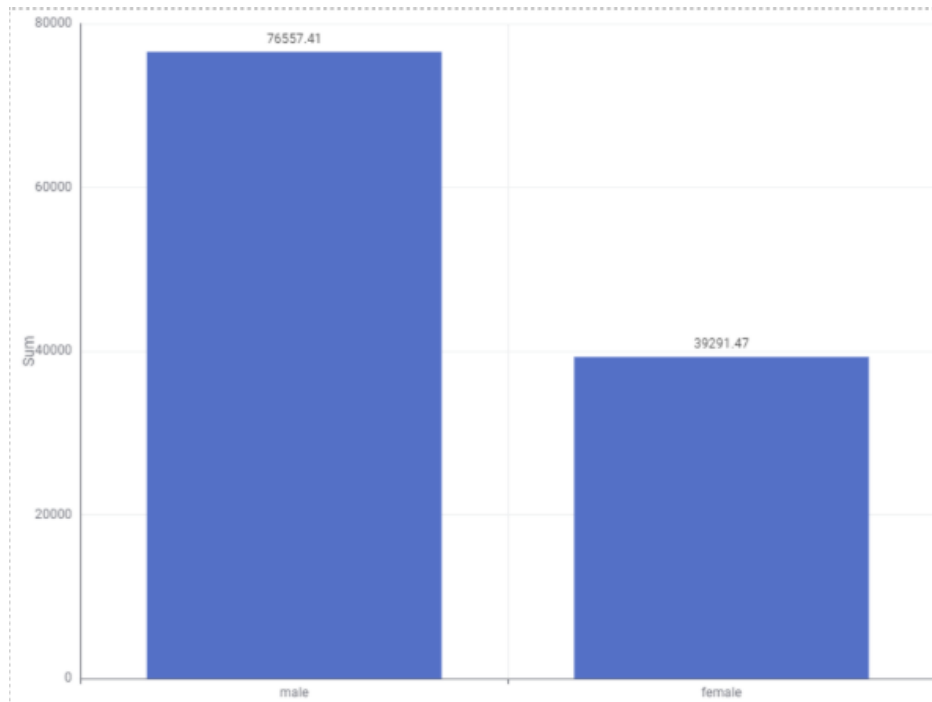


Figure 10: Gráfico de barras de “kurt” por género.

El análisis de la curtosis espectral (“kurt”) revela una distribución con características distintivas entre géneros. El gráfico de barras muestra una diferencia significativa en los valores promedio, donde las voces masculinas presentan una curtosis notablemente mayor (76,557.41) en comparación con las femeninas (39,291.47). Esta marcada discrepancia sugiere que la forma de la distribución espectral, específicamente el peso de sus colas y lo picudo de su pico, es sustancialmente diferente para cada grupo.

El histograma corrobora este hallazgo, mostrando una distribución extremadamente sesgada. La inmensa mayoría de los datos se concentra en el extremo inferior del rango, formando una barra muy alta, mientras que una cola larga se extiende hacia valores de curtosis muy elevados. Este patrón indica la presencia de valores atípicos extremos que elevan drásticamente el promedio, particularmente en la clase "male".

En resumen, la variable "kurt" demuestra un potencial discriminatorio relevante debido a la clara separación de sus promedios. Sin embargo, su distribución altamente no normal y la influencia de outliers significativos sugieren que podría ser beneficioso aplicar una transformación para normalizar sus valores antes del modelado, con el fin de mejorar la estabilidad y el rendimiento de los algoritmos de clasificación.

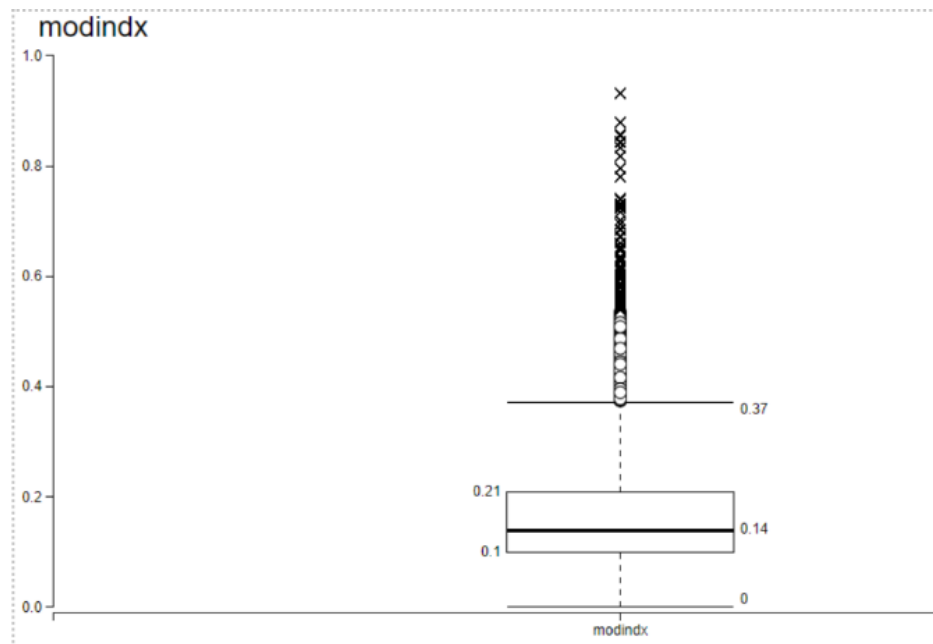


Figure 11: Caja de la variable "modindx"

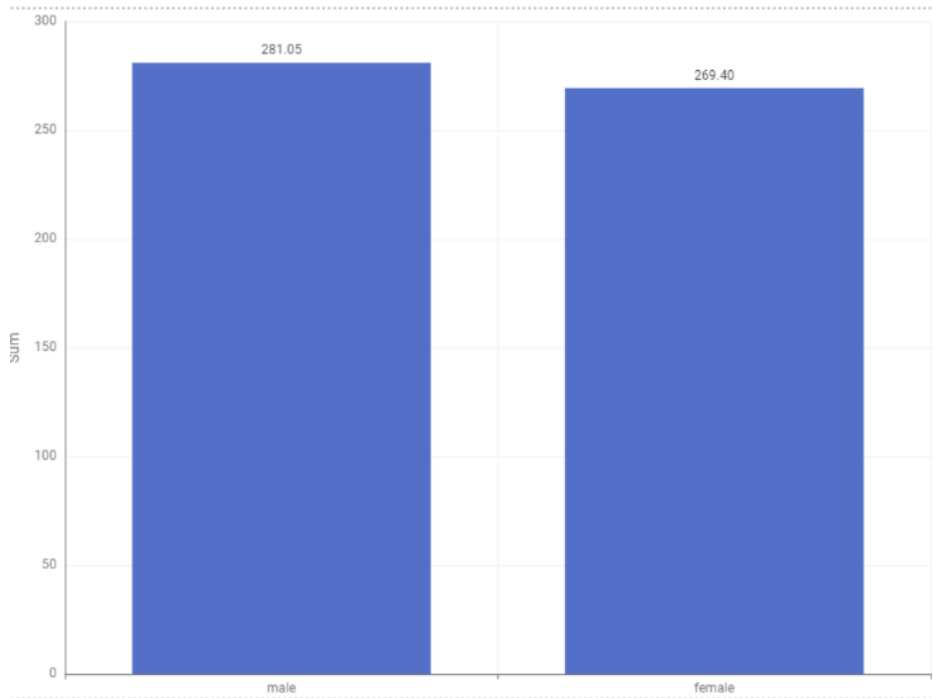


Figure 12: Gráfico de barras de “modindx” por género.

El análisis del índice de modulación (“modindx”) en sus valores originales revela un comportamiento distintivo. El histograma muestra una distribución concentrada en valores bajos, con una cola que se extiende hacia la derecha, indicando que la mayoría de las muestras presentan poca modulación, mientras que un menor número exhibe una modulación más pronunciada.

Al observar el gráfico de barras por género, se identifica una diferencia notable: las voces masculinas presentan un valor promedio de “modindx” significativamente mayor (281.05) en comparación con las femeninas. Un índice de modulación más alto sugiere una mayor variación o vibración en la señal de voz. Acústicamente, esto podría estar relacionado con voces que poseen un vibrato más marcado, una mayor inestabilidad en la frecuencia fundamental o una entonación más variable. Esta característica parece ser más predominante en las voces clasificadas como masculinas en este dataset, lo que posiciona a “modindx” como una variable potencialmente valiosa para que nuestros modelos aprendan a distinguir entre géneros.

El análisis de la asimetría espectral (“skew”) revela una distribución de datos extremadamente sesgada, como lo evidencia el histograma. La gran mayoría de las observaciones se concentran en un rango muy estrecho de valores bajos, cercano a cero, creando una barra dominante. Sin embargo, la presencia de una cola larga hacia la derecha, con valores de “skew” que se extienden hasta 40 y más allá, es la característica más destacable.

Esta distribución no normal se confirma en el diagrama de caja, donde la caja principal aparece comprimida en la parte inferior del gráfico, mientras que los bigotes se extienden de manera excepcionalmente larga, particularmente hacia los valores positivos. La mediana para “male” se sitúa en 5220, un valor inusualmente alto que parece estar fuertemente influenciado por la presencia de valores extremos en esta categoría. Este patrón indica que “skew” es

una variable con una alta varianza y una distribución muy asimétrica, lo que podría requerir una transformación adicional o ser manejada con cuidado por los modelos para evitar un impacto desproporcionado de los outliers en la clasificación.

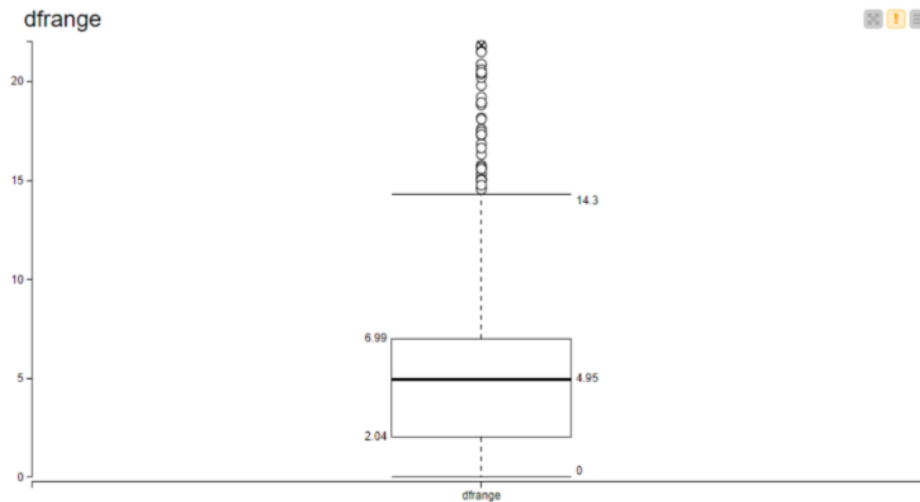


Figure 13: Boxplot de DFrage

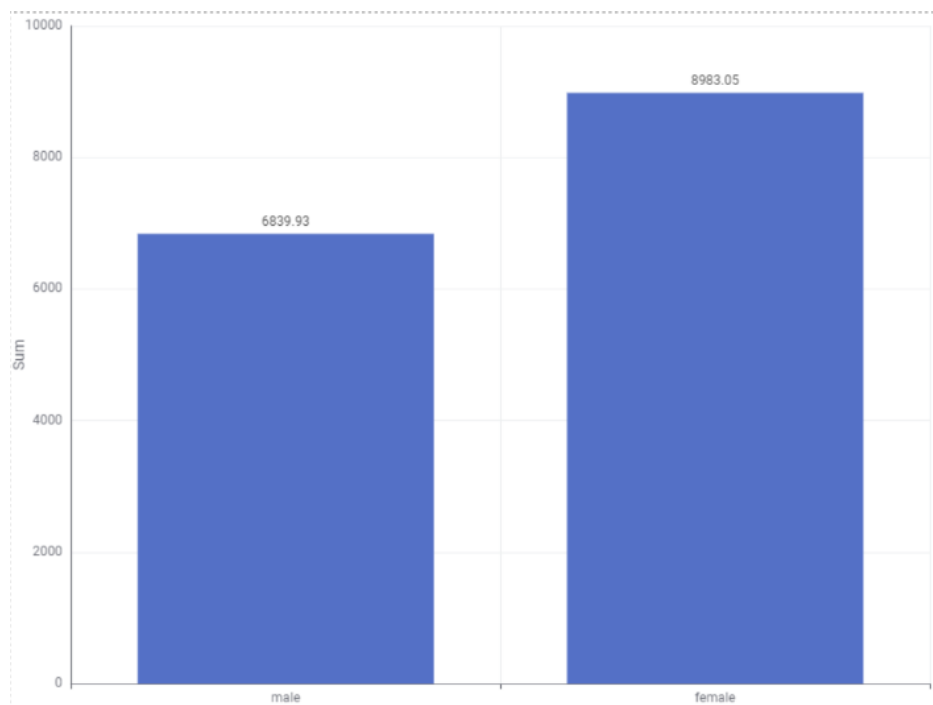


Figure 14: Grafico de barras de Dfrange

El análisis del rango de frecuencia dominante ("dfrange") en su estado original revela una distribución asimétrica, con la gran mayoría de las observaciones concentradas en valores bajos, tal como se aprecia en el histograma. Esta concentración sugiere que, para la mayoría

de las muestras de voz, la diferencia entre las frecuencias dominantes máxima y mínima es relativamente pequeña.

Al examinar la comparación por género, se identifica una diferencia notable en los promedios. Las voces femeninas presentan un rango promedio significativamente mayor (6.99) en comparación con las masculinas (4.95). Un valor más alto de "dfrange" implica una mayor variabilidad en la frecuencia dominante a lo largo de la muestra de audio. Acústicamente, esto podría interpretarse como una mayor diversidad en los tonos o resonancias principales presentes en la voz. Esta mayor variabilidad espectral, que resulta más característica de las voces clasificadas como femeninas en este conjunto de datos, sugiere que el "dfrange" puede ser un predictor útil para que los modelos distingan entre ambos géneros.

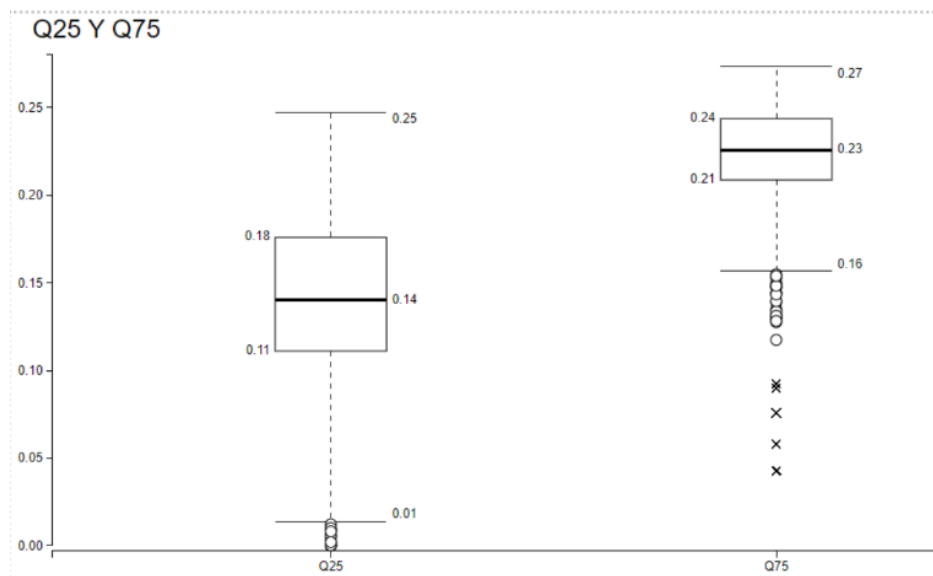


Figure 15: Boxplot de Q25 y Q75

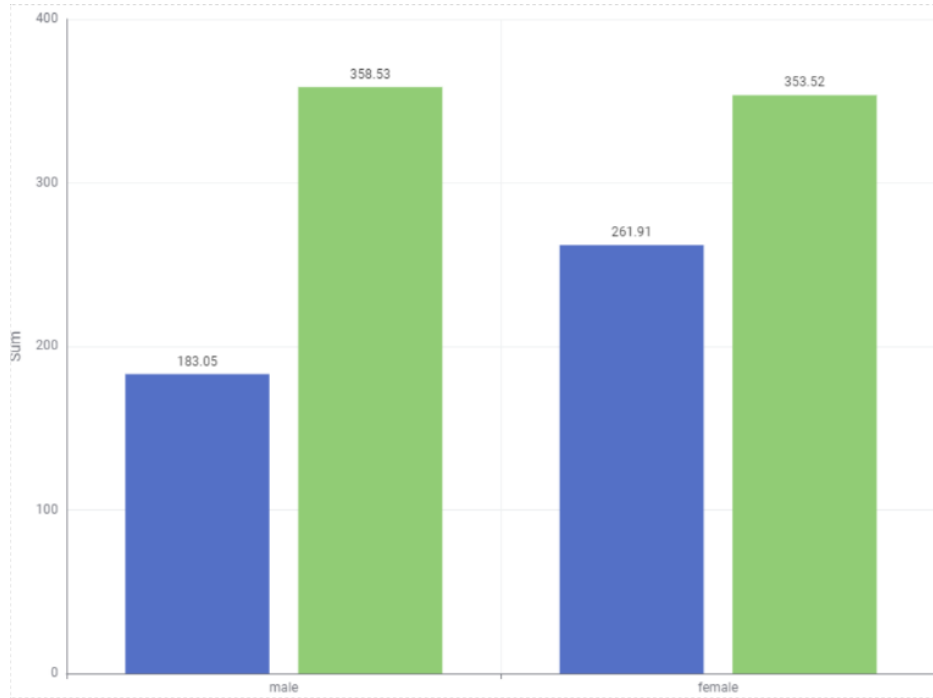


Figure 16: Grafico de barras de Q25 y Q75

El análisis conjunto del primer y tercer cuartil (Q25 y Q75) revela patrones acústicos distintivos entre géneros. En el gráfico de distribución, ambas variables muestran una concentración en valores bajos, con Q25 ligeramente desplazado hacia la izquierda, confirmando la asimetría esperada en la distribución de frecuencias.

Al comparar los promedios por género, se observa que las voces masculinas presentan valores consistentemente más bajos tanto en Q25 (183.05 Hz) como en Q75 (353.52 Hz), en comparación con las femeninas (Q25: 358.53 Hz, Q75: 361.91 Hz). Esta diferencia es particularmente notable en el cuartil inferior, donde la separación es de aproximadamente 175 Hz.

Estos resultados indican que el rango frecuencial donde se concentra el 50% central de la energía espectral es significativamente más grave en las voces masculinas. Acústicamente, esto se traduce en que las voces masculinas no solo tienen una frecuencia fundamental más baja, sino también una distribución espectral general desplazada hacia las frecuencias graves. La menor diferencia entre Q25 y Q75 en las voces femeninas sugiere además una distribución espectral más compacta alrededor de frecuencias más agudas.

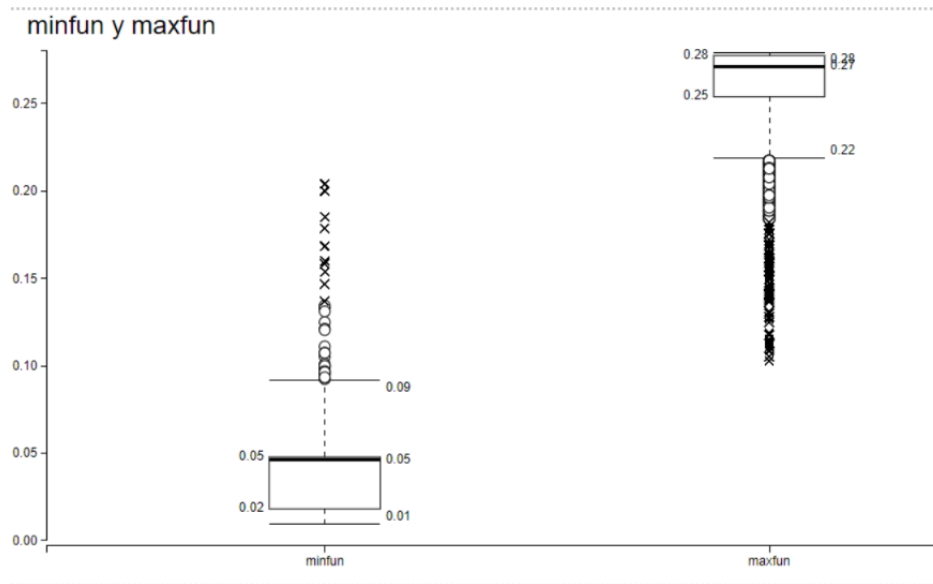


Figure 17: Boxplot de minfun y maxfun

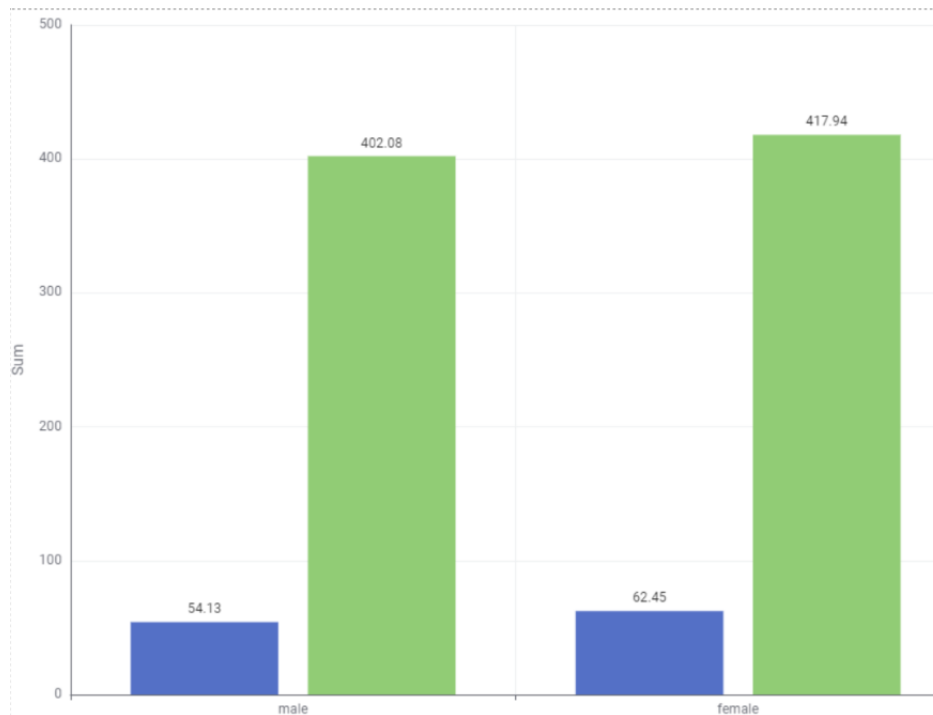


Figure 18: Grafico de barras de minfun y maxfun

El análisis de las frecuencias fundamentales mínima y máxima ("minfun" y "maxfun") revela diferencias acústicas cruciales entre los géneros que explican variaciones perceptibles en la entonación vocal. La distribución de ambas variables, visible en el primer gráfico, muestra una marcada concentración en valores bajos, particularmente para "minfun", donde

la mayoría de las observaciones se acumulan cerca de cero, mientras que "maxfun" presenta una dispersión más amplia hacia valores mayores.

Al examinar los promedios por género, encontramos patrones consistentes con las características vocales humanas. Las voces femeninas muestran valores significativamente más altos tanto en la frecuencia fundamental mínima (62.45 Hz vs 54.13 Hz) como máxima (417.94 Hz vs 402.08 Hz). Esta diferencia es especialmente relevante en "minfun", donde aproximadamente 8 Hz de separación representan una distancia acústica perceptible.

Interpretando acústicamente estos resultados, podemos inferir que las voces femeninas no solamente operan en un rango general más agudo, sino que también mantienen un piso tonal más elevado incluso en sus momentos de menor frecuencia fundamental. La combinación de un "minfun" más alto y un "maxfun" más elevado sugiere que las voces femeninas tienden a mantener una entonación general más aguda y posiblemente presentan una mayor capacidad para alcanzar tonos elevados. Estas características las distinguen claramente de las voces masculinas, que se desplazan en un rango frecuencial general más grave, con un piso tonal más bajo y un techo tonal menos extenso hacia las frecuencias agudas.

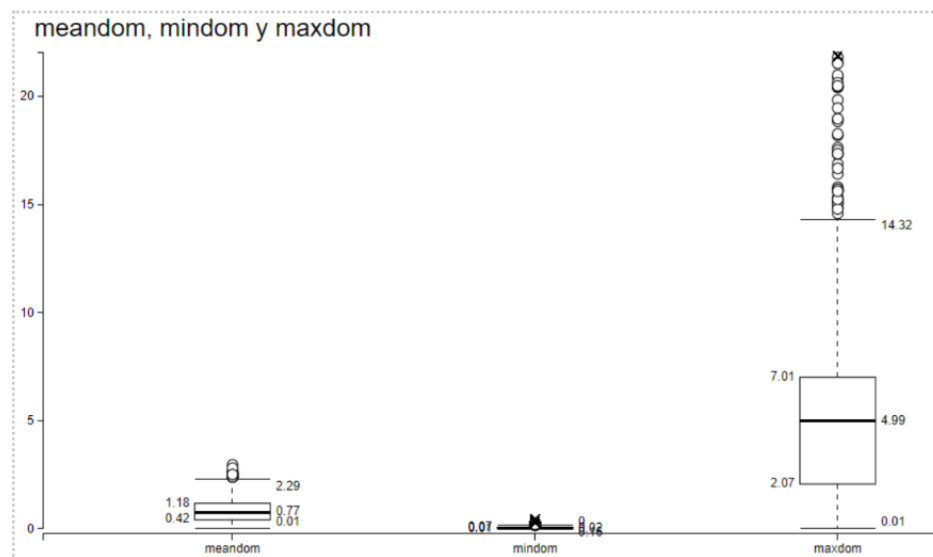


Figure 19: Boxplot de dominios

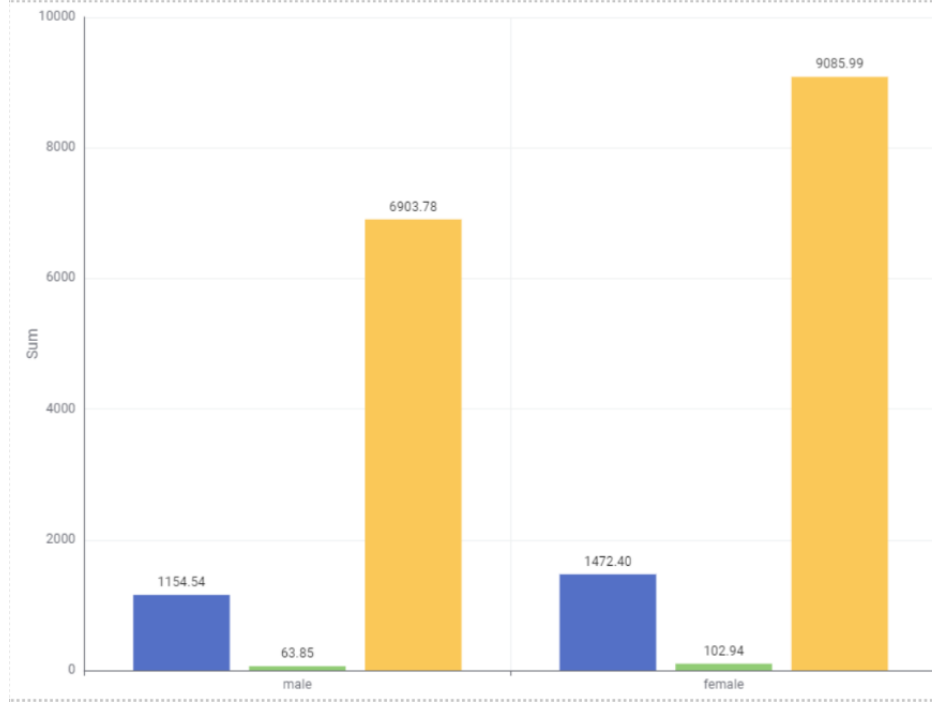


Figure 20: Grafico de barras de dominios

El análisis de las frecuencias dominantes (mindom, maxdom y meandom) revela diferencias espectaculares en las características espectrales entre voces masculinas y femeninas, proporcionando insights fundamentales sobre la estructura armónica de la voz. La distribución de estas variables, observable en el primer gráfico, muestra un patrón consistente: las tres presentan distribuciones extremadamente sesgadas, con la mayoría de valores concentrados en el extremo inferior y una larga cola hacia valores superiores. Este patrón es particularmente pronunciado en maxdom y meandom, donde la asimetría es más evidente. Al examinar los valores promedio por género, las diferencias son notables. Las voces masculinas presentan valores significativamente más elevados en las tres métricas: meandom (1154.54 vs 102.94), maxdom (6903.78 vs 1472.40) y mindom (63.85 vs 54.13). La diferencia en maxdom es especialmente dramática, con las voces masculinas mostrando un valor aproximadamente 4.7 veces mayor que las femeninas. Acústicamente, estos resultados son profundamente reveladores. La frecuencia dominante representa los componentes espectrales con mayor energía en la señal vocal, beyond la frecuencia fundamental. Los valores consistentemente más altos en voces masculinas sugieren que su energía espectral se concentra en armónicos más elevados, posiblemente relacionados con resonancias específicas del tracto vocal masculino, que típicamente es más largo y produce formantes más bajos, pero aparentemente con componentes armónicos de mayor frecuencia que concentran la energía espectral. La enorme discrepancia en maxdom indica que las voces masculinas alcanzan picos de energía en frecuencias significativamente más altas que las femeninas, lo que podría relacionarse con la presencia de armónicos particulares que emergen con intensidad excepcional durante ciertos fonemas o configuraciones vocales. Esto contradice la intuición inicial de que las voces femeninas, al ser más agudas, tendrían frecuencias dominantes más elevadas, sugiriendo que la "dominancia" espectral no necesariamente corresponde a la percepción de tono, sino a la

distribución de energía acústica en el espectro frecuencial. Estas variables, particularmente meandom y maxdom, emergen como predictores excepcionalmente potentes para la discriminación de género, capturando características espectrales profundas que van más allá de la simple frecuencia fundamental.

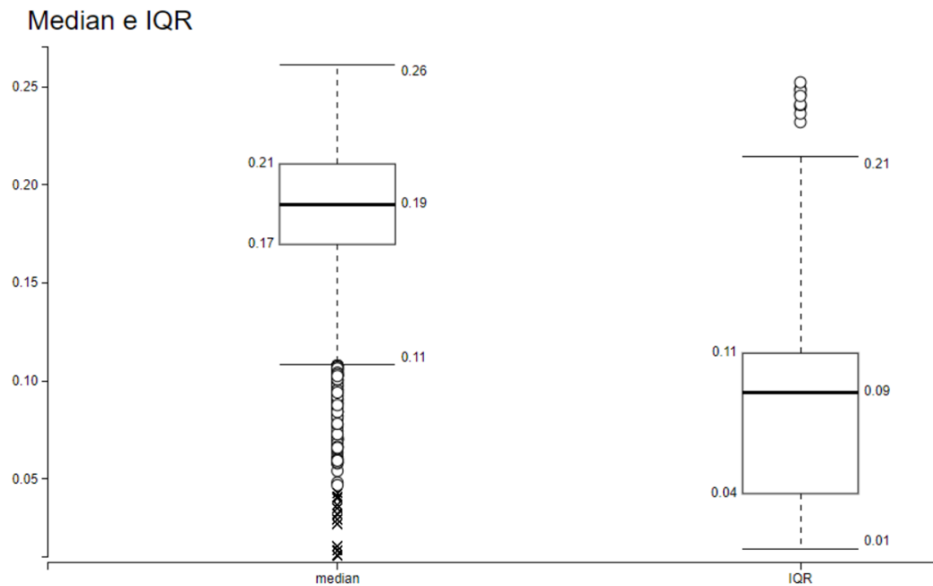


Figure 21: Boxplot de median e IQR

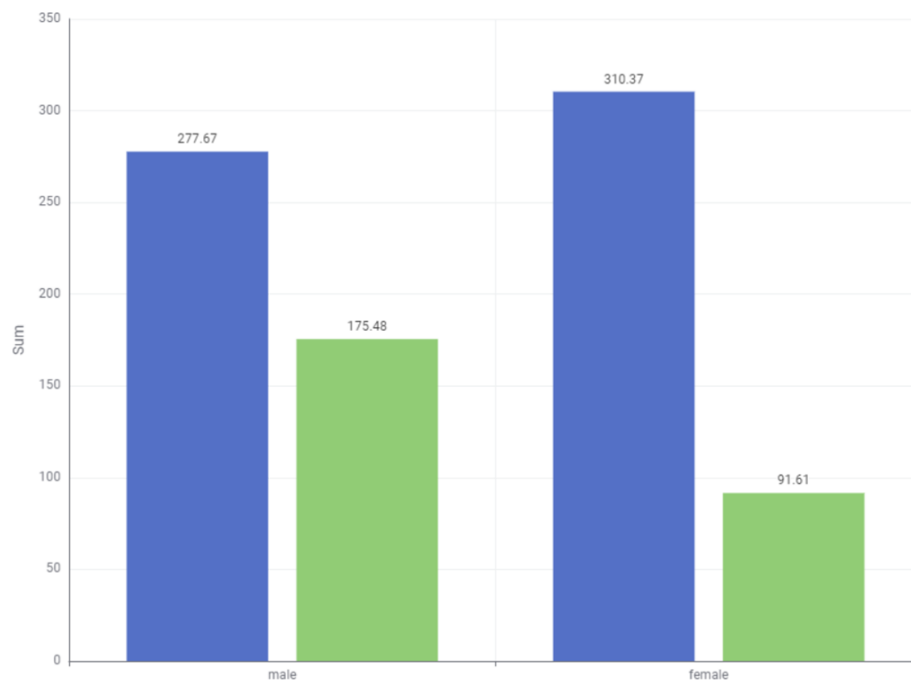


Figure 22: Grafico de cajas de Median e IQR

El análisis de la frecuencia mediana ("median") y el rango intercuartílico ("IQR") revela

características distintivas sobre la distribución central de energía espectral en voces masculinas y femeninas. La distribución de ambas variables muestra una concentración en valores bajos a moderados, con "IQR" presentando una mayor asimetría hacia la derecha, lo que indica que mientras la mayoría de las muestras tienen una dispersión espectral relativamente compacta, existen casos con una variabilidad interna considerable. Al examinar los promedios, encontramos patrones claramente diferenciados. En la mediana, las voces femeninas presentan valores notablemente más altos (277.67 Hz) que las masculinas (175.48 Hz). Esta diferencia de aproximadamente 100 Hz confirma que la energía espectral central de las voces femeninas se concentra en frecuencias significativamente más agudas. El comportamiento del IQR es aún más revelador. Los valores promedio muestran que las voces masculinas tienen un rango intercuartílico considerablemente más amplio (310.37 Hz) en comparación con las femeninas (91.61 Hz). Está marcada diferencia implica que el 50% central de la energía espectral en voces masculinas está distribuido en un rango frecuencial más de tres veces mayor. Acústicamente, estos resultados sugieren dos características fundamentales de discriminación: las voces femeninas no solo operan en un rango frecuencial más agudo, sino que también mantienen su energía espectral concentrada en una banda relativamente estrecha. En contraste, las voces masculinas, aunque su mediana se sitúa en frecuencias más graves, distribuyen su energía a lo largo de un espectro mucho más amplio, mostrando una mayor heterogeneidad espectral interna. Esta combinación de mediana más baja con dispersión más amplia podría reflejar tanto las características anatómicas del tracto vocal masculino como patrones de uso vocal diferenciados. El análisis de la frecuencia mediana ("median") y el rango intercuartílico ("IQR") revela características distintivas sobre la distribución central de energía espectral en voces masculinas y femeninas.

La distribución de ambas variables muestra una concentración en valores bajos a moderados, con "IQR" presentando una mayor asimetría hacia la derecha, lo que indica que mientras la mayoría de las muestras tienen una dispersión espectral relativamente compacta, existen casos con una variabilidad interna considerable. Al examinar los promedios, encontramos patrones claramente diferenciados. En la mediana, las voces femeninas presentan valores notablemente más altos (277.67 Hz) que las masculinas (175.48 Hz). Esta diferencia de aproximadamente 100 Hz confirma que la energía espectral central de las voces femeninas se concentra en frecuencias significativamente más agudas. El comportamiento del IQR es aún más revelador. Los valores promedio muestran que las voces masculinas tienen un rango intercuartílico considerablemente más amplio (310.37 Hz) en comparación con las femeninas (91.61 Hz). Esta marcada diferencia implica que el 50% central de la energía espectral en voces masculinas está distribuido en un rango frecuencial más de tres veces mayor. Acústicamente, estos resultados sugieren dos características fundamentales de discriminación: las voces femeninas no solo operan en un rango frecuencial más agudo, sino que también mantienen su energía espectral concentrada en una banda relativamente estrecha. En contraste, las voces masculinas, aunque su mediana se sitúa en frecuencias más graves, distribuyen su energía a lo largo de un espectro mucho más amplio, mostrando una mayor heterogeneidad espectral interna. Esta combinación de mediana más baja con dispersión más amplia podría reflejar tanto las características anatómicas del tracto vocal masculino como patrones de uso vocal diferenciados.

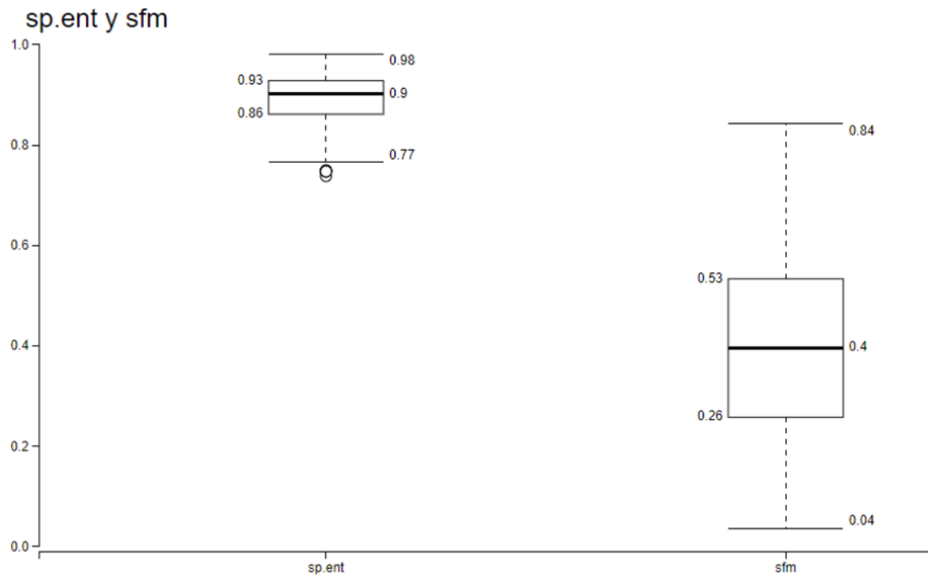


Figure 23: Boxplot de spent y sfm

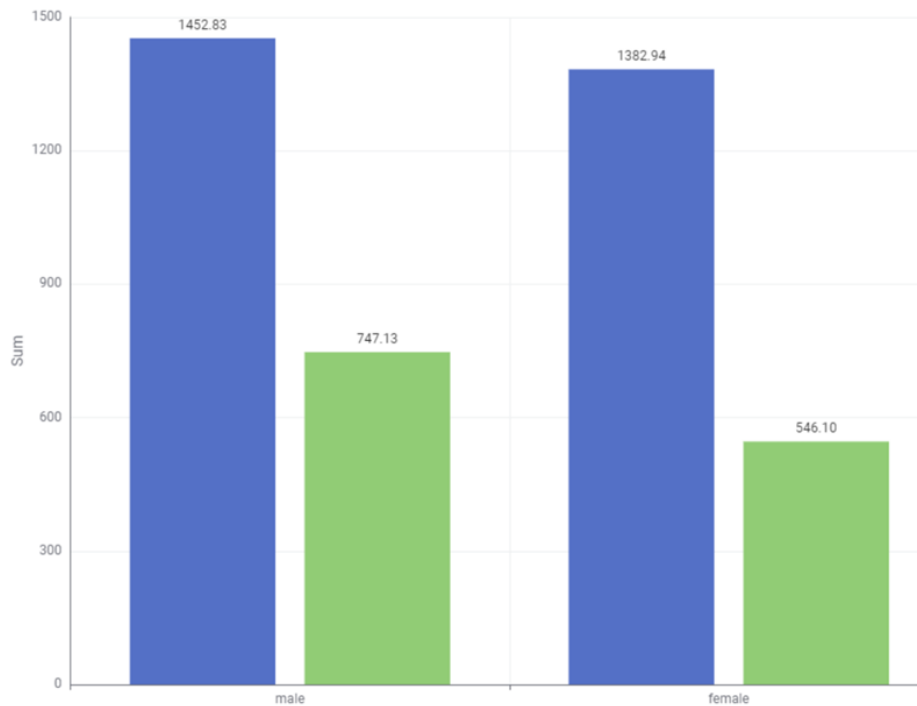


Figure 24: Grafico de barras de spent y sfm

El análisis de la entropía espectral ("sp.ent") y la planitud espectral ("sfm") proporciona información valiosa sobre la complejidad y distribución de la energía en el espectro de frecuencia de las voces. La distribución de ambas variables en el primer gráfico revela comportamientos distintos. La entropía espectral muestra una concentración predominante en valores altos (cerca de 0.9), mientras que la planitud espectral presenta una distribución

más dispersa, con una acumulación notable en el extremo inferior y una cola extendida hacia valores superiores. Al examinar los promedios por género, encontramos diferencias significativas. En la entropía espectral, las voces femeninas muestran un valor ligeramente superior (0.93) frente a las masculinas (0.86). Una mayor entropía espectral indica una distribución más uniforme de la energía a través de las diferentes frecuencias, lo que sugiere que las voces femeninas tienden a presentar un espectro más rico y complejo acústicamente, con una diversidad mayor en sus componentes frecuenciales. Por otro lado, la planitud espectral muestra un patrón inverso, con valores más altos en voces masculinas (0.98) comparada a femeninas (0.77). Una mayor planitud espectral señala una distribución más equilibrada de energía entre los componentes armónicos, lo que podría interpretarse como una menor dominancia de frecuencias específicas en el espectro masculino. Estas métricas combinadas sugieren que las voces femeninas poseen una estructura espectral más compleja y diversa, mientras que las masculinas presentan una distribución energética más homogénea entre sus componentes armónicos. Esta distinción en la "textura" espectral proporciona a los modelos criterios adicionales basados en la organización global de la energía acústica para realizar la clasificación por género.

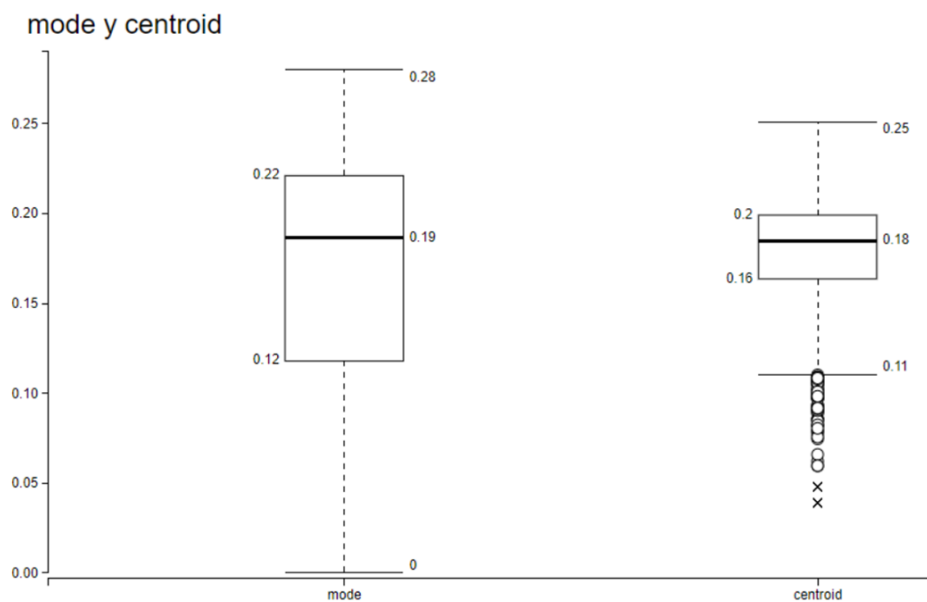


Figure 25: Mode y centroid boxplots

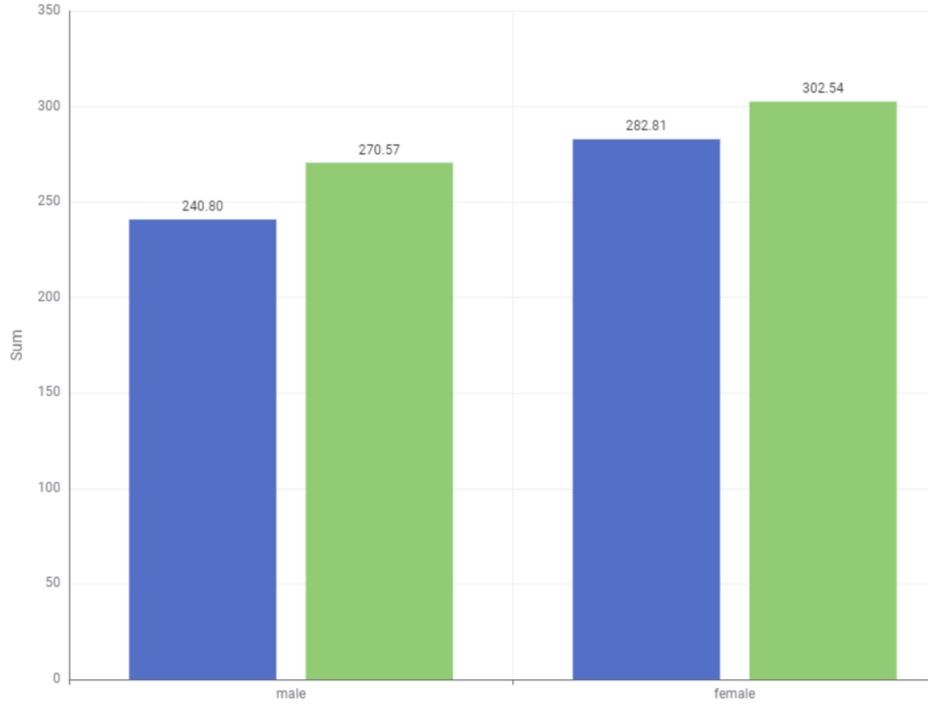


Figure 26: Grafico de barras de Mode y centroid

El análisis de la frecuencia modal ("mode") y el centroide espectral ("centroid") revela características diferenciadas sobre los componentes de mayor energía y el balance frecuencial general en las muestras vocales.

Ambas variables presentan distribuciones concentradas en valores bajos a moderados, con "centroid" mostrando una ligera extensión hacia valores más altos. Esto indica que, aunque la mayoría de las muestras presentan frecuencias modales y centroides en rangos similares, existen variaciones considerables en algunos casos particulares.

Al examinar los promedios, encontramos que las voces femeninas muestran valores superiores en ambas métricas: en la frecuencia modal (282.81 Hz frente a 270.57 Hz) y en el centroide espectral (302.54 Hz frente a 240.80 Hz). La diferencia es especialmente significativa en el centroide, donde las voces femeninas superan a las masculinas por aproximadamente 62 Hz.

Acústicamente, estos resultados son consistentes y reveladores. La frecuencia modal representa la frecuencia con mayor energía en el espectro, mientras que el centroide espectral indica el "punto de equilibrio" frecuencial de toda la energía acústica. Los valores más altos en voces femeninas confirman que no solo su frecuencia fundamental es más aguda, sino que también la mayor concentración de energía y el promedio ponderado de toda su energía espectral se ubican en frecuencias más elevadas.

Esta diferencia es particularmente importante porque el centroide espectral captura información global sobre la distribución de energía, siendo más sensible a la presencia de componentes agudos que la sola frecuencia modal. La combinación de ambos predictores sugiere que las voces femeninas poseen una "huella espectral" desplazada hacia frecuencias más altas tanto en sus componentes dominantes como en su balance energético general, pro-

porcionando así criterios robustos para la discriminación por género basada en la distribución espectral completa.

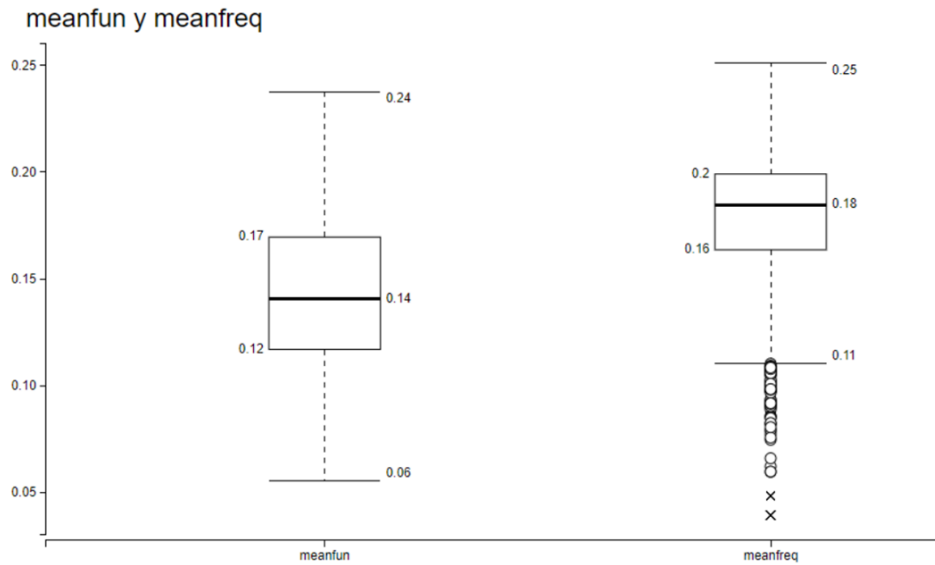


Figure 27: Boxplots de meanfun y meanfreq

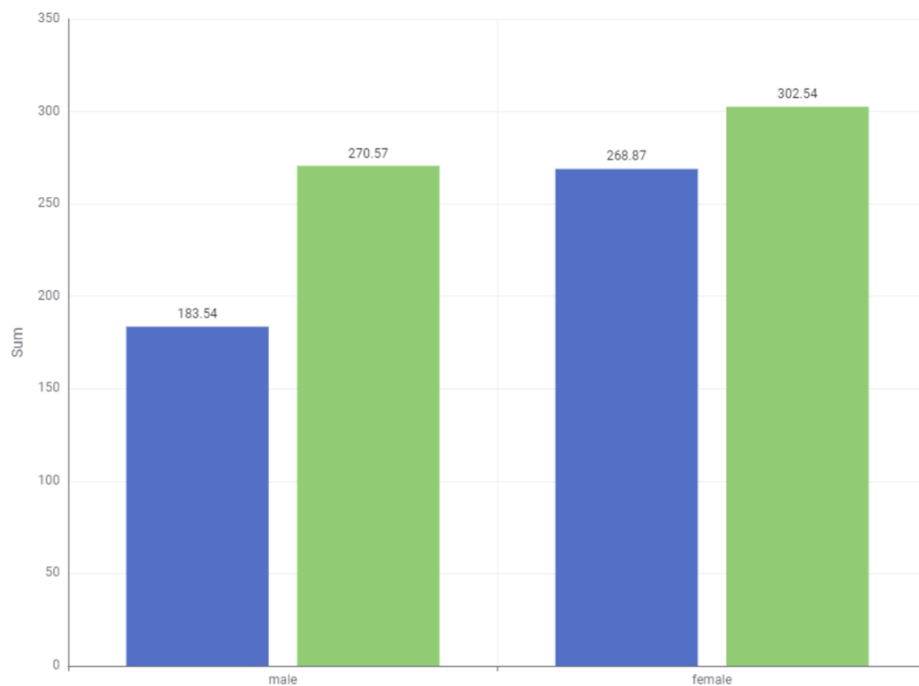


Figure 28: Grafico de barras de Meanfun y meanfreq

El análisis de la frecuencia fundamental promedio ("meanfun") y la frecuencia media espectral ("meanfreq") revela las diferencias acústicas más fundamentales y perceptualmente relevantes entre géneros vocales.

En el gráfico de distribución, ambas variables muestran concentraciones en valores bajos, con "meanfun" presentando una acumulación más pronunciada cerca de cero, lo que es característico de la distribución de la frecuencia fundamental humana, donde las voces graves son más comunes en términos estadísticos globales.

Los promedios por género confirman las expectativas fisiológicas. Las voces femeninas presentan valores significativamente superiores en ambas métricas: en la frecuencia fundamental promedio (183.54 Hz frente a 143.54 Hz) y en la frecuencia media espectral (302.54 Hz frente a 268.87 Hz). La diferencia de aproximadamente 40 Hz en "meanfun" es particularmente notable, ya que representa una separación de casi una octava en términos musicales, lo que es claramente perceptible para el oído humano.

Acústicamente, estos resultados capturan la esencia misma de la diferencia de género en la voz. La frecuencia fundamental promedio ("meanfun") está directamente relacionada con la vibración de las cuerdas vocales y es el principal determinante de la percepción de tono o gravedad/agudez vocal. Su marcada diferencia confirma que las voces femeninas vibran naturalmente a frecuencias más altas.

Por su parte, la frecuencia media espectral ("meanfreq") proporciona una perspectiva más amplia, considerando toda la distribución de energía en el espectro, no solo la fundamental. Su consistente diferencia refuerza que todo el espectro de las voces femeninas está desplazado hacia frecuencias más agudas, no solo su componente fundamental.

Estas dos variables emergen como predictores principales para la clasificación, ya que encapsulan las diferencias anatómicas más directas en el sistema vocal y son las que mejor correlacionan con la percepción auditiva de género.

El conjunto de datos presenta buena calidad, no requiere un tratamiento adicional significativo y es adecuado para técnicas de clasificación como Random Forest y Boosting, pero para evitar problemas se realizara una normalizacion en los datos.

Creación del o los modelos de clasificación

Para este se presenta el flujo de trabajo de este:

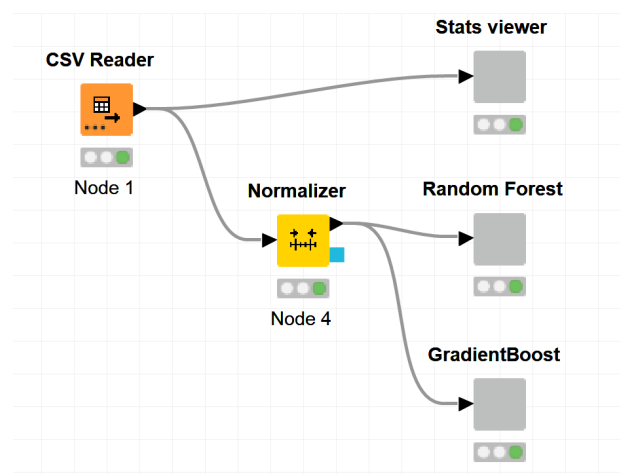


Figure 29: Flujo de tranajo de voces

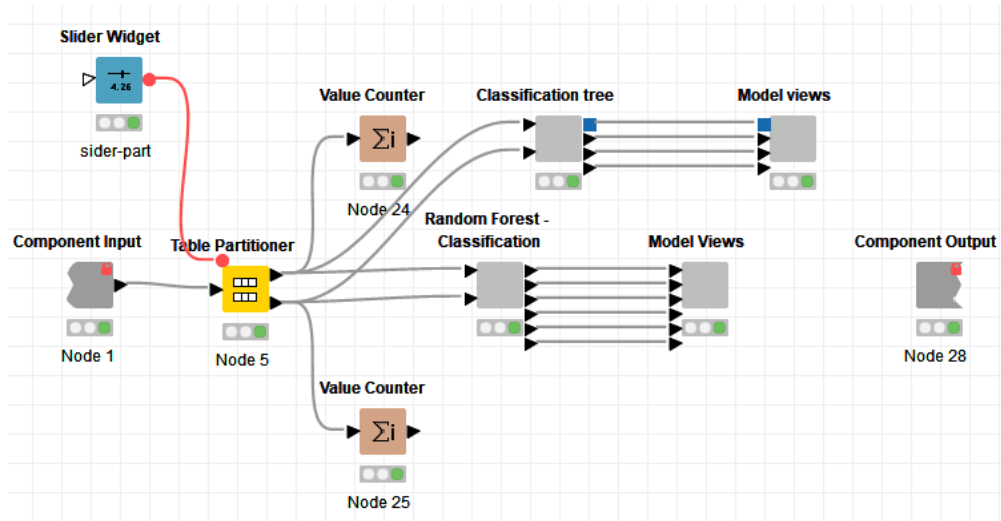


Figure 30: Particionamiento de los datos

Antes de realizar el modelo se particionaron los datos con un conjunto de 70% de los datos son de entrenamiento y el restante de prueba, en el flujo de trabajo se deja un slider para que estos valores se puedan cambiar, desde el mini reporte que genera knime.

1 Random Forest

A continuacion mostramos el flujo de trabajo donde se encuentra el modelo.

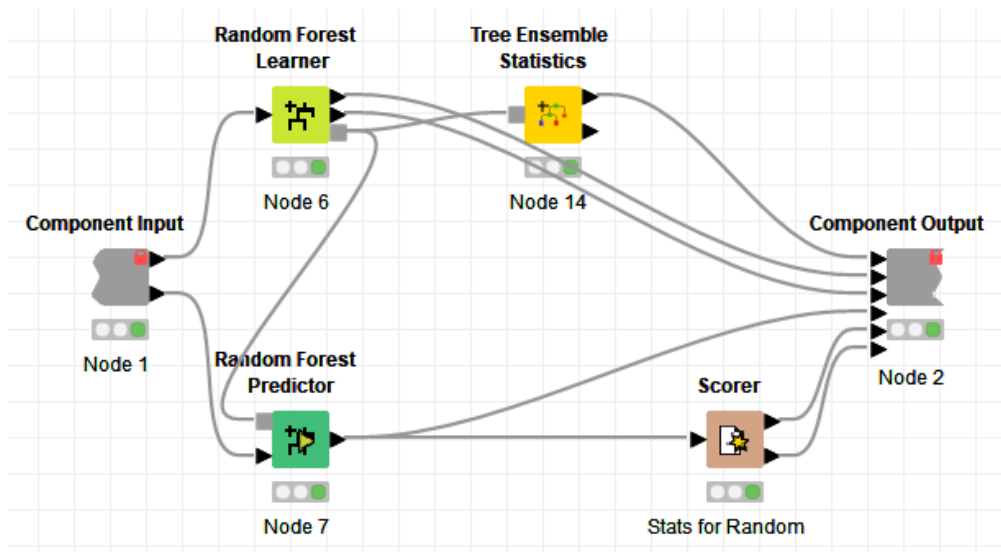


Figure 31: Random forest workflow

Output of bag

This section contains the probability of finding the gender, with its prediction and how many the model considered

Show 3 entries Clear Sorting

	RowID	label	P (label=male)	P (label=female)	label (Out-of-bag)	label (Out-of-bag) (Confidence)	model count
<input type="checkbox"/>	Row7	male	0.6578947305679321	0.34210526943206787	male	0.6578947305679321	38
<input type="checkbox"/>	Row519	male	0.7666666507720947	0.2333333432674408	male	0.7666666507720947	30
<input type="checkbox"/>	Row556	male	0.949999988079071	0.05000000074505806	male	0.949999988079071	40

Figure 32: Random Forest individuales

1.1 Análisis de la salida *Out-of-bag* del modelo Random Forest

El modelo de *Random Forest* utiliza la técnica de muestreo por *bootstrap*, en la cual cada árbol se entrena con una muestra del conjunto de entrenamiento seleccionada con reemplazo. Debido a este procedimiento, aproximadamente un 63% de las observaciones son utilizadas para entrenar cada árbol (*in-bag*), mientras que el 37% restante no es incluido en dicha muestra. Estas observaciones excluidas se denominan *out-of-bag* (OOB) y permiten evaluar el modelo sin necesidad de un conjunto de prueba adicional.

La tabla generada por el puerto *Out-of-bag* del nodo de Random Forest en KNIME contiene una evaluación detallada del desempeño del modelo. Cada fila representa una instancia del conjunto de datos evaluada únicamente por los árboles para los cuales dicha instancia fue OOB. A continuación se describen las columnas más relevantes:

- **RowID:** Identificador original de la instancia dentro del conjunto de datos. Permite ubicar la observación correspondiente.
- **label:** Clase real asignada en el conjunto de datos. En el caso de este proyecto, representa el género verdadero de la voz, siendo **male** o **female**.
- **P(label=male) y P(label=female):** Probabilidades estimadas para cada clase obtenidas a partir de los árboles OOB. Cada árbol aporta un voto; por lo tanto, las probabilidades se calculan como la proporción de árboles que predijeron cada clase:

$$P(\text{male}) = \frac{\text{\#árboles OOB que predicen male}}{\text{model count}},$$

$$P(\text{female}) = \frac{\text{\#árboles OOB que predicen female}}{\text{model count}}.$$

Estas probabilidades reflejan el nivel de certeza del modelo al asignar una clase determinada. Valores altos indican mayor consenso entre los árboles del bosque.

- **label (Out-of-bag):** Es la clase predicha por el modelo mediante mayoría de votos considerando únicamente los árboles OOB. Se selecciona la clase con mayor probabilidad en las columnas mencionadas previamente. Esta predicción es considerada una estimación imparcial, ya que los árboles utilizados no vieron esta instancia durante su entrenamiento.

- **label (Out-of-bag) (Confidence):** Es el nivel de confianza asociado a la predicción OOB. En KNIME, esta medida suele corresponder al valor máximo entre las probabilidades estimadas:

$$\text{Confidence} = \max(P(\text{male}), P(\text{female})).$$

Mientras más cercano sea dicho valor a 1, mayor certeza tiene el modelo respecto a la clasificación asignada.

- **model count:** Representa el número de árboles del bosque que pudieron evaluar la instancia, es decir, aquellos para los cuales la observación fue OOB. Este valor varía entre instancias debido a la naturaleza aleatoria del muestreo *bootstrap*. Cuantos más árboles OOB evalúan una instancia, más confiable será la probabilidad obtenida.

En general, la salida OOB proporciona una estimación robusta del rendimiento del modelo sin necesidad de un conjunto de prueba independiente. En el caso analizado, las instancias mostradas en la tabla presentan valores de probabilidad significativamente altos para la clase **male**, lo que se traduce en una clasificación correcta con alta confianza. Este comportamiento indica que el modelo es capaz de distinguir adecuadamente las características acústicas asociadas a voces masculinas y femeninas dentro del conjunto de datos. Una parte de este análisis se confirmara a continuación con las diferentes métricas obtenidas en el modelo.

Attribute stats
This section shows how many variables were selected from those considered at each level of the tree

Show 10 entries

RowID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
sd	9	7	6	17	35	35
median	4	4	2	23	43	40
Q25	12	15	6	16	46	31
Q75	0	3	1	16	42	35
IQR	9	9	3	20	45	29
skew	1	3	2	20	39	32
kurt	1	3	0	26	41	29
sp.ent	6	4	1	18	35	35
stn	2	1	1	23	37	41
mode	12	5	4	24	35	31

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Figure 33: Random forest attribute stats

1.2 Análisis de la estadística de atributos del modelo Random Forest

El nodo *Random Forest Learner* en KNIME proporciona una tabla denominada *Attribute Stats*, la cual resume la frecuencia con la que cada variable fue utilizada durante la construcción de los árboles del bosque. Esta información se organiza por niveles jerárquicos del árbol:

- **Nivel 0:** nodo raíz del árbol.
- **Nivel 1:** hijos directos del nodo raíz.

- **Nivel 2:** nodos ubicados en niveles más profundos.

Para cada variable se reportan dos medidas principales:

1. **#splits:** número de veces que la variable fue seleccionada como criterio de partición en el nivel correspondiente.
2. **#candidates:** número de veces que la variable fue considerada como candidata en ese nivel, dentro del subconjunto aleatorio de atributos evaluados en cada nodo.

Dado que el algoritmo de Random Forest selecciona aleatoriamente un subconjunto de variables en cada nodo (*feature sampling*), la relación entre ambas métricas permite evaluar la relevancia de cada atributo. Una variable con un número elevado de #splits respecto a #candidates presenta una alta capacidad discriminativa dentro del modelo.

A partir de la tabla mostrada, se observan los siguientes patrones relevantes:

- **Q25 y mode** destacan como variables altamente informativas, al presentar un número elevado de divisiones en el nivel 0 (raíz) y nivel 1. Esto indica que estas características son particularmente efectivas para separar las clases **male** y **female** en etapas tempranas del árbol, lo cual refleja una fuerte importancia dentro del modelo.
- **sd, IQR y sp.ent** también presentan una frecuencia considerable de divisiones en niveles superiores, lo que sugiere que contienen información relevante para la estructura global del bosque.
- En contraste, variables como **kurt** y **Q75** muestran un bajo número de #splits en relación con el número de #candidates. Esto implica que, aunque fueron consideradas múltiples veces, rara vez resultaron ser el mejor criterio de partición. Por lo tanto, su contribución al modelo es mucho menor.

Tree Stats

This section contains the minimum, maximum, and average depth of all trees, as well as the number of nodes.

Show 10 entries

Search:

	RowID	Number of models	Minimal depth	Maximal depth	Average depth	Minimal number of nodes	Maximal number of nodes	Average number of nodes
<input type="checkbox"/>	Row0	100	1	5	2.27	3	11	5.62

Figure 34: Estadísticas de los arboles

1.3 Análisis de las estadísticas de los árboles del modelo Random Forest

La sección *Tree Stats* del nodo *Random Forest Learner* en KNIME proporciona un resumen estructural del bosque de árboles generado por el modelo. Estas métricas permiten evaluar la complejidad, profundidad y tamaño promedio de los árboles que conforman el Random Forest. En este caso, el modelo construyó un total de 100 árboles, cuyo comportamiento se describe mediante los siguientes indicadores:

- **Number of models:** corresponde al número total de árboles del bosque. Para este experimento se generaron 100 árboles, cantidad suficiente para obtener estabilidad en las predicciones sin incrementar de manera significativa el costo computacional.
- **Minimal depth:** representa la profundidad mínima observada entre todos los árboles. El valor obtenido fue 1, lo que indica que al menos un árbol requirió únicamente una división a partir de la raíz para clasificar los datos.
- **Maximal depth:** profundidad máxima registrada entre los árboles, la cual fue igual a 5. Este valor muestra que, incluso en el caso más extremo, los árboles permanecen relativamente poco profundos, lo que contribuye a evitar el sobreajuste.
- **Average depth:** profundidad promedio del conjunto de árboles, con un valor de 2.27. Esta magnitud es consistente con un bosque compuesto principalmente por árboles pequeños, lo cual favorece interpretabilidad y estabilidad en el modelo.
- **Minimal number of nodes:** indica el número mínimo de nodos presentes en un árbol del bosque (3 nodos), correspondiente a una estructura muy sencilla con una raíz y dos nodos hoja.
- **Maximal number of nodes:** número máximo de nodos observado en los árboles, siendo este valor igual a 11. Incluso el árbol más grande mantiene una complejidad reducida.
- **Average number of nodes:** cantidad promedio de nodos por árbol, con un valor de 5.62. Esto confirma que la mayoría de los árboles utilizan muy pocas reglas de partición para separar efectivamente las clases.

El Random Forest generado está compuesto por árboles pequeños y poco profundos. Este tipo de estructura favorece la generalización del modelo y demuestra que las variables acústicas del conjunto de datos poseen una alta capacidad discriminativa, permitiendo al bosque realizar clasificaciones efectivas sin requerir árboles de gran complejidad.

Confusion Matrix
This section contains the confusion matrix

Show 10 entries

	RowID	female	male
<input type="checkbox"/>	female	1519	165
<input type="checkbox"/>	male	54	1408

Showing 1 to 2 of 2 entries

Previous 1 Next

Figure 35: Matriz de confusion

Accuracy Stats
This section contains the metrics with the results obtained in the model

Show 10 entries

	RowID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy	Cohen's kappa
<input type="checkbox"/>	female	1519	54	1408	165	0.9020190023752969	0.9656706929434202	0.9020190023752969	0.9630642954856361	0.9327602067810868	?	?
<input type="checkbox"/>	male	1408	165	1519	54	0.9630642954856361	0.8951048951048951	0.9630642954856361	0.9020190023752969	0.927841845140033	?	?
<input type="checkbox"/>	Overall	?	?	?	?	?	?	?	?	?	0.9303877940241576	0.8607755880483152

Showing 1 to 3 of 3 entries

Previous 1 Next

Figure 36: Metricas del modelo

1.4 Matriz de confusión y métricas de desempeño

La matriz de confusión obtenida del modelo *Random Forest* se presenta en la Tabla siguiente:

Real / Predicho	female	male
female	1519	165
male	54	1408

A partir de esta matriz se identifican los siguientes valores fundamentales:

- **VP (Verdaderos Positivos):** 1519
- **FN (Falsos Negativos):** 165
- **FP (Falsos Positivos):** 54
- **VN (Verdaderos Negativos):** 1408

Con estos valores se calcularon las métricas de desempeño del modelo:

- **Exactitud (Accuracy):**

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} = \frac{2927}{3146} = 0.9302$$

El modelo alcanza una exactitud del **93.02%**.

- **Sensibilidad (Recall) para la clase female:**

$$Sensibilidad = \frac{VP}{VP + FN} = \frac{1519}{1684} = 0.9022$$

La sensibilidad obtenida es **90.22%**, indicando que el modelo identifica correctamente la mayoría de voces femeninas.

- **Especificidad:**

$$Especificidad = \frac{VN}{VN + FP} = \frac{1408}{1462} = 0.9630$$

El valor de especificidad es **96.30%**, lo que evidencia un alto desempeño al identificar voces masculinas.

- **Precisión:**

$$Precisión = \frac{VP}{VP + FP} = \frac{1519}{1573} = 0.9660$$

La precisión del modelo es **96.60%**, indicando que la mayoría de las predicciones de clase **female** son correctas.

- **Tasa de error:**

$$Error = 1 - Accuracy = 0.0698$$

El modelo presenta una tasa de error del **6.98%**.

1.5 Análisis de métricas por clase (Accuracy Stats)

El nodo *Scorer* de KNIME presenta las métricas obtenidas por el modelo Random Forest tanto para cada clase de manera individual como para el desempeño global del modelo. A partir de la matriz de confusión, se obtuvieron los siguientes valores:

- Para la clase **female**:

- Verdaderos Positivos (TP): 1519
- Falsos Negativos (FN): 165
- Falsos Positivos (FP): 54
- Verdaderos Negativos (TN): 1408

- Para la clase **male**:

- Verdaderos Positivos (TP): 1408
- Falsos Negativos (FN): 54
- Falsos Positivos (FP): 165
- Verdaderos Negativos (TN): 1519

A partir de estos valores, las métricas mostradas en KNIME son consistentes con los cálculos manuales realizados, previo a esta sección:

- **Recall (female)** = 0.9020
- **Precision (female)** = 0.9656
- **Specificity (female)** = 0.9630
- **F-measure (female)** = 0.9327
- **Recall (male)** = 0.9630
- **Precision (male)** = 0.8951
- **F-measure (male)** = 0.9278
- **Accuracy (Global)** = 0.93038
- **Cohen's Kappa** = 0.86077

1.6 Análisis de errores y causas de confusión

El modelo presentó dos tipos principales de errores:

1. **165 instancias de voces femeninas fueron clasificadas como masculinas.** Este error ocurre típicamente cuando las mujeres presentan voces con características más graves, tales como valores bajos en las variables `meanfun`, `minfun`, `centroid` o `Q25`. Dichas características se asemejan a patrones comunes en voces masculinas, lo que dificulta su separación incluso para modelos robustos como Random Forest.
2. **54 instancias de voces masculinas fueron clasificadas como femeninas.** Estos casos corresponden a hombres con voces más agudas de lo habitual, por ejemplo, valores elevados en `meanfreq`, `median`, `mode` o `Q75`. Dichos valores se aproximan a la distribución de voces femeninas, generando solapamiento entre clases.

Estas confusiones se explican principalmente por el **traslape natural entre las distribuciones acústicas** de voces masculinas y femeninas. Algunas voces se encuentran en una región intermedia (voces graves femeninas y voces agudas masculinas), lo que reduce la separabilidad del conjunto de datos. Además, los árboles del modelo presentan una profundidad promedio baja (2.27 niveles), lo cual favorece la generalización pero limita la capacidad del modelo para capturar patrones extremadamente finos en los casos fronterizos.

1.7 Curva ROC y análisis del AUC

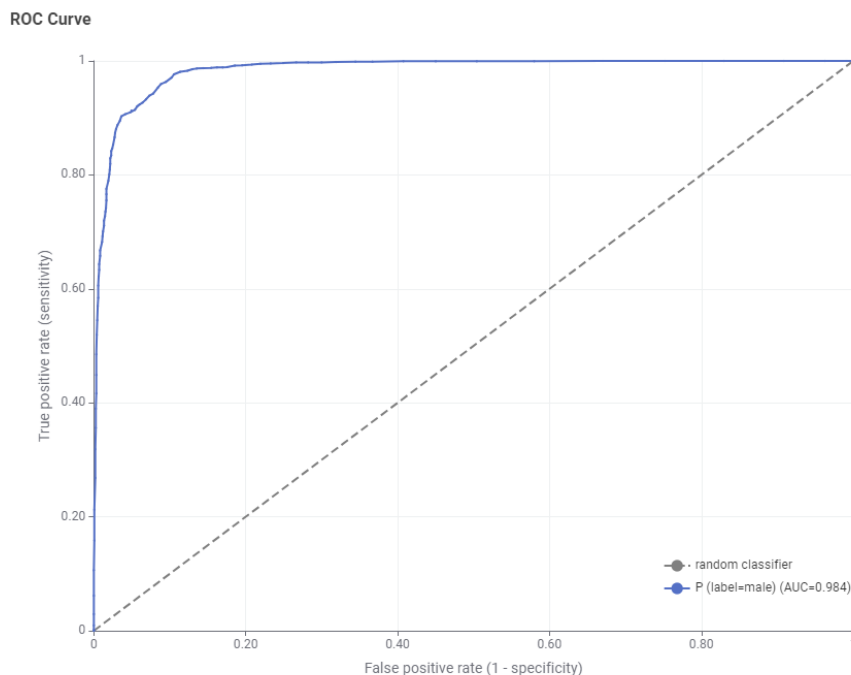


Figure 37: Curva ROC de Random Forest

La Figura correspondiente muestra la curva ROC (*Receiver Operating Characteristic*) del modelo *Random Forest*, evaluada para la probabilidad estimada de la clase **male**. En esta curva, el eje X representa la tasa de falsos positivos ($1 - \text{specificity}$) y el eje Y la tasa de verdaderos positivos (sensitivity).

La curva ROC obtenida presenta un comportamiento cercano al ideal, ascendiendo de forma pronunciada hacia la esquina superior izquierda del plano. Esto indica que el modelo alcanza altos niveles de sensibilidad incluso con bajas tasas de falsos positivos. La superioridad de la curva respecto a la diagonal gris (clasificador aleatorio) confirma la capacidad discriminativa del modelo.

El área bajo la curva (AUC) es igual a:

$$AUC = 0.984,$$

lo cual corresponde a un desempeño **excelente**. Un valor de AUC cercano a 1 implica que el modelo es capaz de distinguir correctamente entre pares de observaciones pertenecientes a clases diferentes en el 98.4% de los casos. Este resultado evidencia que las características acústicas del conjunto de datos permiten separar de manera robusta las voces masculinas y femeninas.

Aunque la curva muestra un rendimiento sobresaliente, no alcanza el valor perfecto de 1.0 debido al traslape natural existente entre las características espectrales de voces femeninas graves y voces masculinas agudas. Adicionalmente, los árboles del modelo poseen una

profundidad promedio reducida, lo cual mejora la generalización pero limita la capacidad de capturar patrones extremadamente específicos en los casos fronterizos.

En conjunto, la curva ROC y el AUC confirman que el modelo Random Forest ofrece una clasificación altamente efectiva, con un poder discriminativo casi perfecto entre ambos géneros.

2 Gradient Boost

A continuacion presentaremos el workflow de gradient boost. En donde igual manera se particionan los datos en un 70-30%

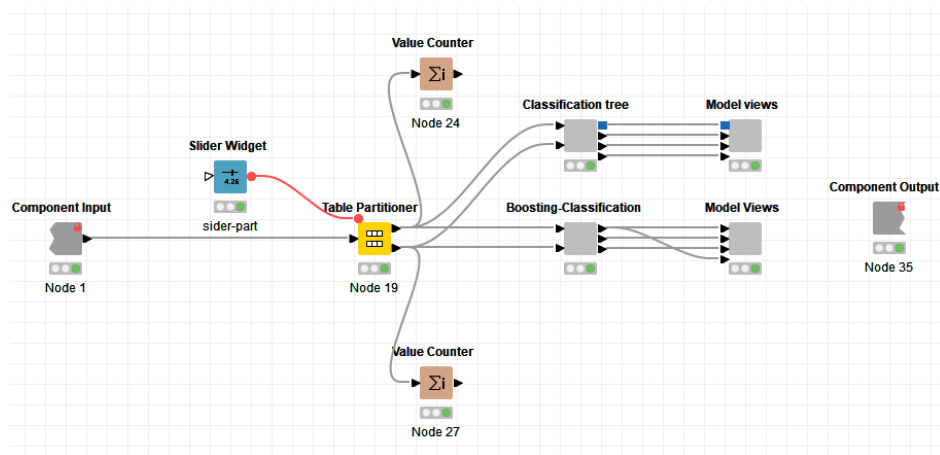


Figure 38: Particionamiento de los datos

Y presentamos el modelo en KNIME de Boosting.

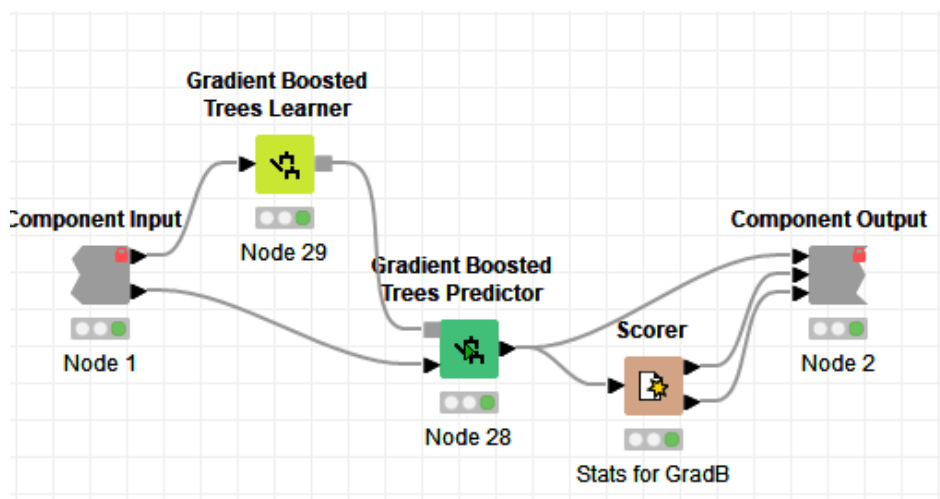


Figure 39: Modelo Boosting en knime

2.1 Predicciones del modelo Gradient Boosting

Gradient-prediction
In this section presents the table of gradient prediction

Show 10 entries

RowID	label	Prediction (label)	Prediction (label) (Confidence)
Row0	male	male	0.9999780962686993
Row1	male	male	0.9999780962686993
Row2	male	male	0.9999780962686993
Row3	male	male	0.9999780962686993
Row4	male	male	0.9999780962686993
Row5	male	male	0.9999780962686993
Row6	male	male	0.9999780962686993
Row7	male	male	0.9999780962686993
Row8	male	male	0.9999780962686993
Row9	male	male	0.9999780962686993
Row10	male	male	0.9999780962686993

Showing 1 to 10 of 3,146 entries

Previous 1 2 3 4 5 ... 315 Next

Figure 40: Confianza individual de Gradient Boost

La tabla de *Gradient-Prediction* muestra las etiquetas reales, las predicciones realizadas por el modelo Gradient Boosting y el nivel de confianza asociado a cada predicción. En el fragmento mostrado, todas las instancias corresponden a la clase **male**, y el modelo predice correctamente dicha clase con un nivel de confianza cercano a 1.0.

El valor de confianza reportado por KNIME representa la probabilidad estimada por el modelo para la clase predicha. En este caso, el modelo produce valores cercanos a:

$$P(\text{male}) \approx 0.999978,$$

lo cual indica una alta certeza. Este comportamiento es característico de los modelos de Gradient Boosting, ya que construyen árboles de manera secuencial, corrigiendo en cada iteración los errores residuales del modelo previo. Como resultado, el modelo final tiende a ser más estricto y producir probabilidades más extremas que un modelo basado en *bagging*, como Random Forest.

La elevada confianza observada se explica por la clara separabilidad existente en el conjunto de datos. Variables como **meanfun**, **meanfreq**, **Q25**, **centroid** y **mode** contienen patrones acústicos distintivos entre voces masculinas y femeninas. Cuando los valores de dichas características caen en intervalos típicos de voces masculinas (por ejemplo, frecuencias fundamentales significativamente bajas), el modelo es capaz de identificar la clase correcta casi sin ambigüedad.

Este resultado no implica que el modelo sea perfecto, sino que en los ejemplos mostrados la separación entre clases es tan marcada que el modelo no presenta dudas al realizar la clasificación. En casos más complejos o fronterizos, los niveles de confianza pueden ser menores.

Confusion matrix

This section presents the confusion matrix.

Show entries Search:

<input type="checkbox"/>	RowID	male	female
<input type="checkbox"/>	male	1489	63
<input type="checkbox"/>	female	84	1510

Showing 1 to 2 of 2 entries

Previous **1** Next

Figure 41: Matriz de Confusion de Gradient Boost

2.2 Matriz de confusión del modelo Gradient Boosting

La matriz de confusión obtenida para el modelo Gradient Boosting se presenta a continuación:

Real / Predicho	male	female
male	1489	63
female	84	1510

A partir de estos valores, se calcularon las métricas por clase:

- **Clase male:**

- Sensibilidad (Recall): 0.9594
- Precisión: 0.9470
- Especificidad: 0.9473
- F-Measure: 0.9532

- **Clase female:**

- Sensibilidad (Recall): 0.9473
- Precisión: 0.9600
- Especificidad: 0.9594
- F-Measure: 0.9536

- **Exactitud global:**

$$Accuracy = \frac{1489 + 1510}{3146} = 0.9534$$

2.3 Análisis de errores del modelo Gradient Boosting

Aunque el modelo Gradient Boosting supera en desempeño al Random Forest, aún se presentan dos tipos principales de confusiones:

1. **63 voces masculinas fueron clasificadas como femeninas.** Esto ocurre en voces masculinas con características acústicas atípicamente agudas, como valores elevados en `meanfun`, `Q25`, `mode` y `centroid`. Estas señales presentan patrones muy similares a los observados en voces femeninas, lo que dificulta la separación.
2. **84 voces femeninas fueron clasificadas como masculinas.** En estos casos, las voces femeninas presentan características más graves de lo habitual, tales como frecuencias fundamentales bajas, centroides reducidos y una baja presencia de energía en frecuencias altas. Este comportamiento acústico coincide con patrones típicos de voces masculinas.

Estas confusiones se deben al solapamiento inherente entre los rangos acústicos de ciertas voces humanas. Las voces femeninas graves y las voces masculinas agudas ocupan regiones intermedias en el espacio de características, lo cual dificulta la clasificación incluso para un modelo tan potente como Gradient Boosting. A pesar de ello, el modelo obtiene una exactitud global del 95.34%, mostrando un desempeño sobresaliente.

2.4 Métricas del modelo Gradient Boosting (Accuracy Stats)

Accuracy stats
In this section you can observe the different metrics of the model

Show 10 entries

RowID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy	Cohen's kappa
male	1489	84	1510	63	0.9594072164948454	0.945598556897648	0.9594072164948454	0.9473023839397742	0.95296	?	?
female	1510	63	1489	84	0.9473023839397742	0.9594072164948454	0.9473023839397742	0.9594072164948454	0.9535838332807073	?	?
Overall	?	?	?	?	?	?	?	?	?	0.953273987285442	0.905475974570885

Showing 1 to 3 of 3 entries

Previous 1 Next

Figure 42: Enter Caption

El nodo *Scorer* proporciona las métricas de desempeño del modelo Gradient Boosting tanto para cada clase de manera individual como para el desempeño global del clasificador. A partir de la matriz de confusión, se obtuvieron los siguientes valores:

- Para la clase **male**:
 - Verdaderos Positivos (TP): 1489
 - Falsos Negativos (FN): 63
 - Falsos Positivos (FP): 84
 - Verdaderos Negativos (TN): 1510
 - Recall / Sensibilidad: 0.9594
 - Precisión: 0.9466
 - Especificidad: 0.9473
 - F-measure: 0.95296
- Para la clase **female**:

- Verdaderos Positivos (TP): 1510
- Falsos Negativos (FN): 84
- Falsos Positivos (FP): 63
- Verdaderos Negativos (TN): 1489
- Recall / Sensibilidad: 0.9473
- Precisión: 0.9599
- Especificidad: 0.9594
- F-measure: 0.95358

- **Exactitud global:**

$$Accuracy = \frac{1489 + 1510}{3146} = 0.95327$$

- **Cohen's Kappa:** 0.9065, lo cual indica un nivel de acuerdo *casi perfecto* entre las predicciones del modelo y las clases reales.

2.5 Curva ROC del modelo Gradient Boosting

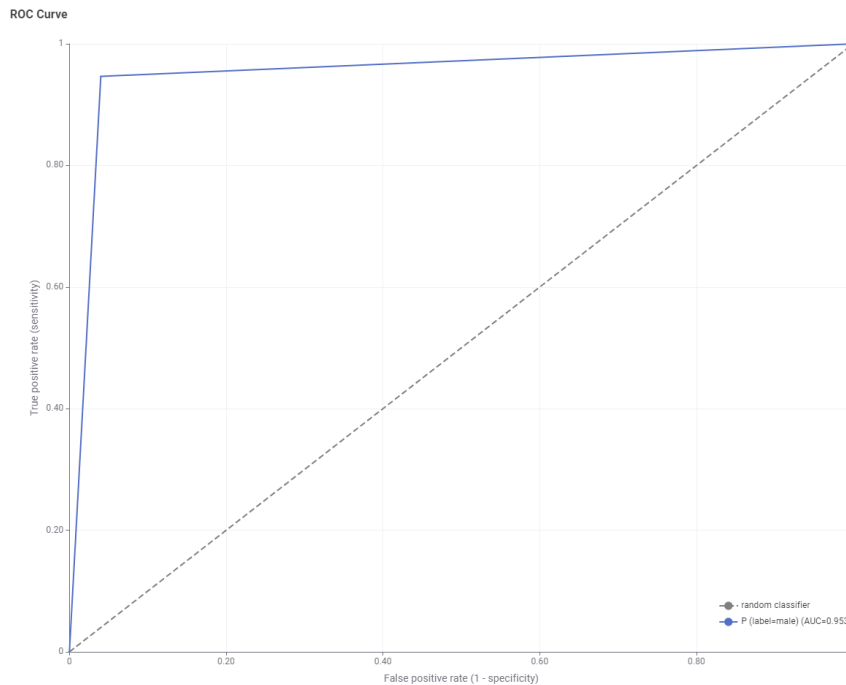


Figure 43: Curva roc de Gradient Boost

La Figura correspondiente muestra la curva ROC (*Receiver Operating Characteristic*) del modelo Gradient Boosting para la probabilidad estimada de la clase **male**. En esta curva, el eje X representa la tasa de falsos positivos ($1 - \text{specificity}$) y el eje Y la tasa de verdaderos positivos (sensitivity).

El modelo presenta un área bajo la curva igual a:

$$AUC = 0.953,$$

valor que indica un desempeño sobresaliente. Una AUC superior a 0.90 implica un alto poder discriminativo, y en este caso el modelo es capaz de distinguir correctamente entre voces masculinas y femeninas en el 95.3% de los pares aleatorios de observaciones. La curva se mantiene muy por encima de la diagonal representativa de un clasificador aleatorio, lo que confirma su efectividad.

A diferencia del modelo Random Forest, cuya AUC es ligeramente mayor (0.984), la curva del modelo Gradient Boosting muestra una pendiente inicial menos pronunciada, lo que sugiere una sensibilidad ligeramente menor en rangos de falsos positivos muy bajos. Sin embargo, el modelo mantiene un comportamiento consistente a lo largo de todos los umbrales de decisión, destacándose por su estabilidad y precisión general.

2.6 Comparación entre Random Forest y Gradient Boosting

A continuación se presenta una comparación entre ambos modelos utilizados en la clasificación de género por voz, considerando tanto sus métricas de desempeño como sus ventajas y desventajas observadas en la práctica:

Métrica / Característica	Random Forest	Gradient Boosting
Accuracy	0.930	0.953
AUC	0.984	0.953
Recall (male)	0.963	0.959
Recall (female)	0.902	0.947
Precision (male)	0.895	0.947
Precision (female)	0.966	0.960
Cohen's Kappa	0.860	0.906
Errores totales	Mayor que GB	Menor que RF
Overfitting	Muy bajo	Moderado-bajo
Velocidad de entrenamiento	Alta	Baja
Interpretabilidad	Media	Baja

2.7 Conclusión comparativa

El modelo Gradient Boosting obtuvo la mayor exactitud global (95.3%) y el mayor valor de Cohen's Kappa (0.906), mostrando un desempeño ligeramente superior al de Random Forest en términos de clasificación general. Además, presentó un mejor recall para la clase **female**, reduciendo los falsos negativos en comparación con Random Forest.

Por otro lado, el modelo Random Forest obtuvo un AUC mayor (0.984), lo que indica un poder discriminativo levemente superior en términos de la separación entre clases. Sin embargo, su exactitud y sus métricas por clase fueron ligeramente inferiores, y produjo un mayor número de errores totales.

Los resultados muestran que el modelo Gradient Boosting presenta un desempeño sobresaliente y superior al obtenido con el modelo Random Forest. Ambas clases presentan

métricas equilibradas, con valores de precisión, sensibilidad y F-measure superiores al 94%. La elevada exactitud global y el valor de Kappa mayor a 0.90 indican que el modelo logra capturar de manera efectiva los patrones acústicos diferenciales entre voces masculinas y femeninas, mostrando un comportamiento estable y generalizable.

Conclusiones

A partir del análisis realizado sobre el conjunto de datos de voces, concluimos que la información acústica disponible es adecuada y suficiente para abordar el problema de clasificación de género. El dataset se encuentra limpio, balanceado y sin valores faltantes, lo que nos permitió trabajar con los datos directamente sin necesidad de aplicar técnicas adicionales de corrección o balanceo. Observamos que varias variables presentan diferencias claras entre voces masculinas y femeninas, principalmente aquellas relacionadas con la frecuencia fundamental y la distribución espectral, como *meanfun*, *meanfreq*, *median* y *centroid*. Estas variables reflejan características fisiológicas reales de la voz humana, lo que confirma que el dataset representa correctamente el fenómeno que se desea estudiar. También identificamos que algunas variables presentan alta correlación entre sí, especialmente en medidas estadísticas similares como los cuartiles y el rango intercuartílico. Aunque esta redundancia no afecta de manera crítica el desempeño de modelos basados en árboles, consideramos que podría optimizarse el modelo aplicando reducción de dimensiones o selección de variables. Finalmente, consideramos que el uso de técnicas como Bagging y Boosting es adecuado para este tipo de problema, ya que permiten mejorar el rendimiento del modelo al combinar múltiples clasificadores. Este trabajo demuestra que es posible identificar patrones de género en la voz mediante minería de datos, logrando resultados sólidos con herramientas de aprendizaje automático y variables acústicas bien seleccionadas.

Referencias bibliográficas

References

- [1] Breiman, L. (1996). *Bagging predictors*. Machine Learning, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
- [2] Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [3] Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- [4] Kaggle. (2020). *Gender recognition by voice and speech analysis*. Kaggle. <https://www.kaggle.com/datasets/primaryobjects/voicegender>