

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

Grupo: _____

No. Equipo: _____

Nombres: Ulises Abdiel Cabello Cardenas

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

Consideraciones encontradas en el conjunto de datos

Describir las consideraciones que encuentre en el conjunto de datos.

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.

Agregue la descripción de las variables dependientes e independientes. Explicando el nombre del algoritmo y sus características.

Tratamiento de datos

Describa cada una de las variables del conjunto de datos, analice el problema que presentan y aplique la técnica de tratamiento de datos según corresponda. Por ejemplo: tipo, valores faltantes, tratamiento necesario a realizar. Describa el proceso desarrollado.

Creación del o los modelos de clasificación

Genere dos modelos de clasificación: Random Forest (Bagging) y Boosting.

Creación de métricas

Analice la matriz de confusión para identificar y explicar los resultados encontrados.

Desarrolle las siguientes métricas y su significado:

- Matriz de confusión
- Sensibilidad
- Precisión
- Tasa de error
- Exactitud
- Especificidad
- Explicación de VP, FP, VN, FN
- Curva ROC: significado e interpretación
- Importancia de atributos dentro del modelo

Análisis de datos

Describa el significado de las métricas generadas a partir del modelo de clasificación.

Identifique dos errores de elementos clasificados erróneamente y describa la razón.

Conclusiones

Agregue las conclusiones de la investigación desarrollada.

Referencias bibliográficas