



# **Pipeline de Ciencia de Datos**

**Piero Emiliano Vizcarra Vargas**

**Orientador:**

**UNSA - Universidad Nacional de San Agustín de Arequipa  
Junio de 2025**

# Índice

<b>1. Preguntas</b>	<b>3</b>
1.1. ¿Qué problemas identifican en el dataset? . . . . .	3
1.2. ¿Qué descubrieron al analizar los datos? . . . . .	4
1.3. ¿Qué reflejan los patrones de tendencia? . . . . .	5
1.4. ¿Cómo varía la morfología promedio de los latidos normales versus los latidos anómalos de las 4 clases mas representativas en las ventanas temporales alrededor de cada anotación? . . . . .	5

# 1. Preguntas

## 1.1. ¿Qué problemas identifican en el dataset?

El dataset de ECG (MIT-BIH Arrhythmia Database) presenta varios problemas notables:

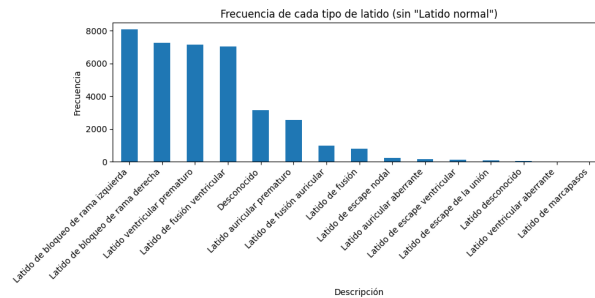


Figura 1: Caption

- **Desbalance de clases:** Existe una distribución muy desigual entre los latidos normales (clase “N”) y las clases anómalas (VEB, SVEB, etc.), lo que puede sesgar los modelos de clasificación.

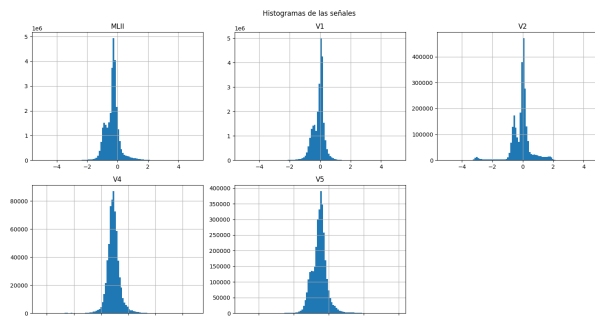


Figura 2: Tendencia Electrodo

- **Variabilidad de derivaciones:** No todos los registros contienen las mismas derivaciones (electrodos), lo que limita el análisis multicanal y la posibilidad de comparar ciertas derivaciones entre pacientes.
- **Datos ruidosos:** Algunos registros pueden contener artefactos o ruido que dificultan la interpretación precisa de las señales.
- **Escasa representación de latidos poco frecuentes:** Clases de arritmias menos comunes tienen muy pocas muestras, lo que complica el entrenamiento de modelos robustos.

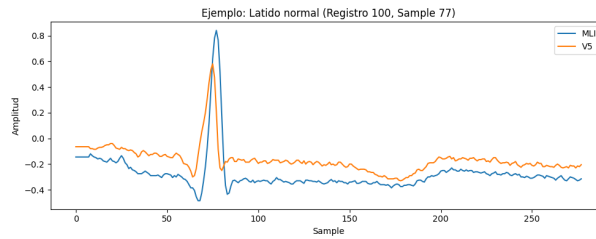


Figura 3: Dato con Ruido

## 1.2. ¿Qué descubrieron al analizar los datos?

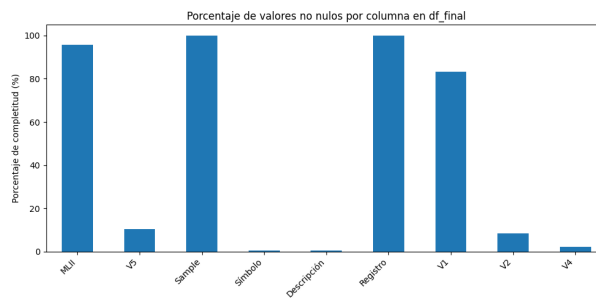


Figura 4: Cantidad de datos completos por columna

- La derivación MLII está presente en casi todos los todos los registros, aun asi lo que la convierte en la señal principal para análisis dado que se puede replicar por otras señales.
- La gran mayoría de los latidos son normales (clase “N”), mientras que las arritmias representan una fracción muy pequeña del total.
- Las anotaciones de latidos anómalos tienden a concentrarse en ciertos segmentos de la señal, mostrando que las arritmias pueden ocurrir de forma intermitente.

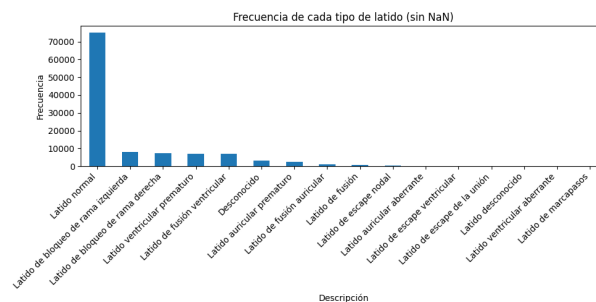


Figura 5: Frecuencia Clasificación con Normal

- Solo hay 2 pacientes que fueron medidos con electrodos V2 y V4, pero no MLII
- Solo un paciente fue medido con el V4 pero si tiene MLII

- Se puede llegar a amputar datos de la V4 dado que no distorcia su pequeño número de datos respecto a la cantidad final
- Los registros contienen información temporal, lo cual sugiere que es crucial tener en cuenta el contexto de la señal para un análisis confiable.
- Bajo el método de energía umbral es posible confundir varias señales normales con etiquetados como normales, así que se debe tener cuidado.

### 1.3. ¿Qué reflejan los patrones de tendencia?

- La señal MLII muestra un ritmo cardíaco regular para la mayoría de los latidos, con ondas P, QRS y T bien definidas.
- Los episodios de arritmias aparecen como interrupciones localizadas en estos patrones regulares, indicando eventos anómalos puntuales.
- La mayoría de los datos presenta estabilidad en las características de las ondas (amplitud y forma), salvo en los casos donde aparecen latidos anómalos.
- Estos patrones permiten identificar ventanas temporales relevantes para un análisis más detallado o para entrenamiento de modelos de clasificación.

### 1.4. ¿Cómo varía la morfología promedio de los latidos normales versus los latidos anómalos de las 4 clases más representativas en las ventanas temporales alrededor de cada anotación?

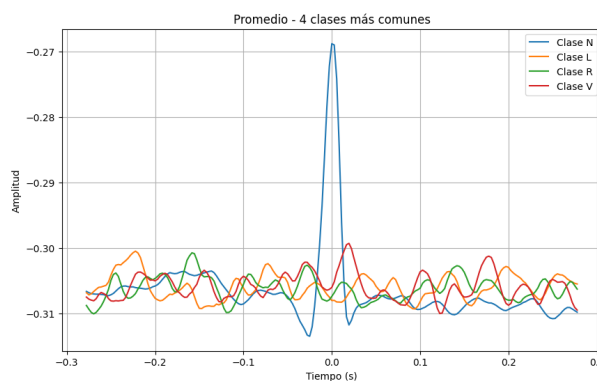


Figura 6: Promedio

El gráfico de promedio revela diferencias claras entre la clase de latido normal (N) y las clases anómalas (L, R y V):

- **Clase N (Normal):** Presenta un pico pronunciado y definido en el punto cero de la escala temporal, característico del complejo QRS de un latido cardíaco normal.

Esta forma estable y simétrica refleja la regularidad de la actividad cardíaca en condiciones normales.

- **Clase L (Bloqueo de rama izquierda):** Muestra un complejo QRS más ancho que la clase normal, indicando una propagación eléctrica más lenta a través de los ventrículos.
- **Clase R (Bloqueo de rama derecha):** Exhibe desviaciones en la forma y amplitud del pico principal, lo que refleja una alteración en la conducción eléctrica.
- **Clase V (Latido ventricular prematuro):** Se observa un ensanchamiento y distorsión significativa en el complejo QRS, con una amplitud reducida en comparación con la clase normal, lo que es característico de los latidos ventriculares prematuros.