

ELECTROCARDIOGRAMA (ECG)

Nadia Martin

Alejandro Wohlwend

Luis Oliari



Definición de problema

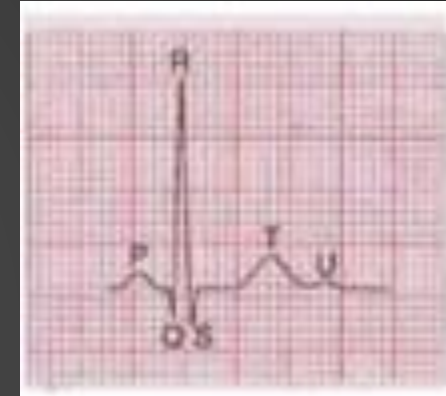
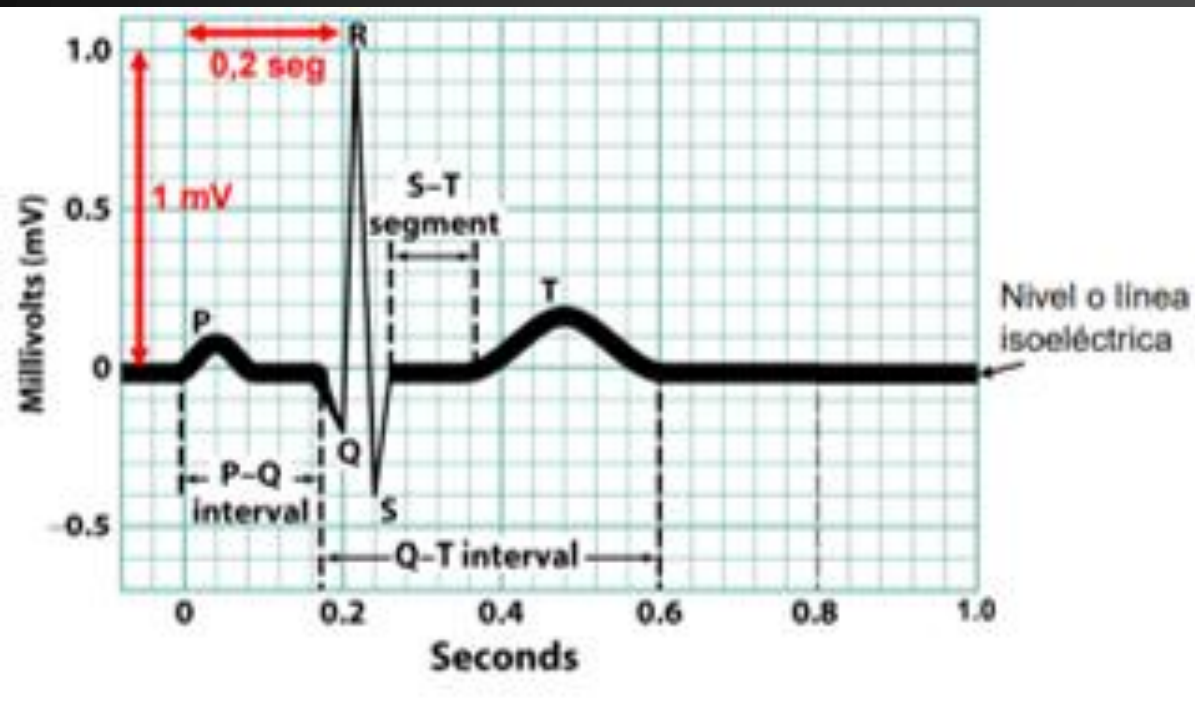
- Predecir diferentes arritmias en base a un electrocardiograma (ECG).

Base de datos

- La base de datos de arritmias de MIT-BIH contiene 48 extractos de media hora de registros de ECG ambulatorios de dos canales, obtenidos de 47 sujetos estudiados por el Laboratorio de Arritmias de BIH entre 1975 y 1979.

fuentes: <https://physionet.org/content/mitdb/1.0.0/>

QUE ES UN ECG?



EDA

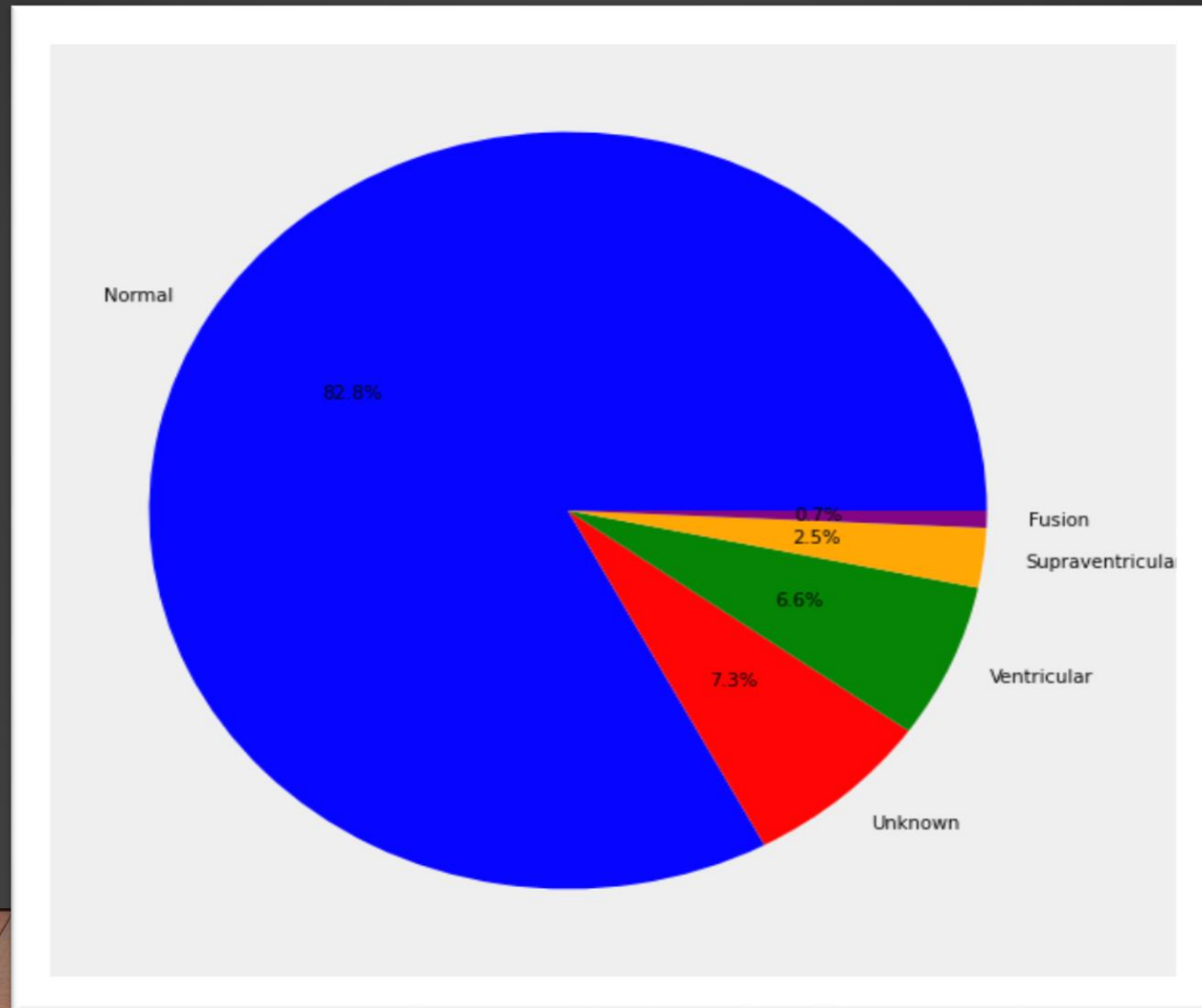
- Un dataset train de 87554 observaciones.
- Un dataset test de 21892 observaciones.
- Una relación de (95:5) entre dataset train y test.
- El dataset tiene 186 variables numéricas independientes y una variable categórica dependiente.
- Las variables están normalizadas entre (0-1)mV y autocorrelacionadas (se registra la lectura de potencia en intervalos de a 0,04seg), tomadas a 125Hz.

- Las clases a predecir son las siguientes.
- Con un conjunto de datos nulos a partir del dato 140.

	Clases
0	Normales
1	Supraventricular
2	Ventricular
3	Fusion
4	Sin clasificar



UN DATASET COMPLETAMENTE DESBALANCEADO:



ALGORITMOS SELECCIONANDOS

- Random Forest
- Boosting
- Redes Neuronales Convoluciones

Estrategia para trabajar con database desbalanceados para la clasificación de clases

- Oversampling (datos sintéticos – SMOTE).
- Subsampling (librería submuestra randomicamente las clases mayoritarias con/sin reposición.
- Combinados.

RANDOM FOREST

	Nº	Dataset	Manejo de desbalance	Recall c/clases					Peor coef. Matriz confusión	Obs.
				0	1	2	3	4		
RF	1	Completo	Class_weight = None (default)	1,00	0,58	0,86	0,60	0,94	0,58	* Semejante matriz confusión
	2	Completo	Class_weight = balanced	1,00	0,58	0,86	0,58	0,94	0,58	* Semejante matriz confusión
	3	Completo	SMOTE	0,99	0,76	0,94	0,77	0,97	0,76	
	4	Completo	Class_weight (RandomSeachCV)	1,00	0,59	0,81	0,59	0,94	0,59	
	5	Completo	Undersampling	0,87	0,81	0,87	0,91	0,95	0,81	** Semejante matriz confusión
	6	Completo	Undersampling (RandomSeachCV)	0,88	0,82	0,88	0,91	0,95	0,82	** Semejante matriz confusión
	7	.iloc[:,0:140]	Class_weight=None (default)	1,00	0,58	0,86	0,60	0,94	0,58	* Semejante matriz confusión
	8	.iloc[:,0:140]	Class_weight=balanced	1,00	0,58	0,86	0,58	0,93	0,58	* Semejante matriz confusión

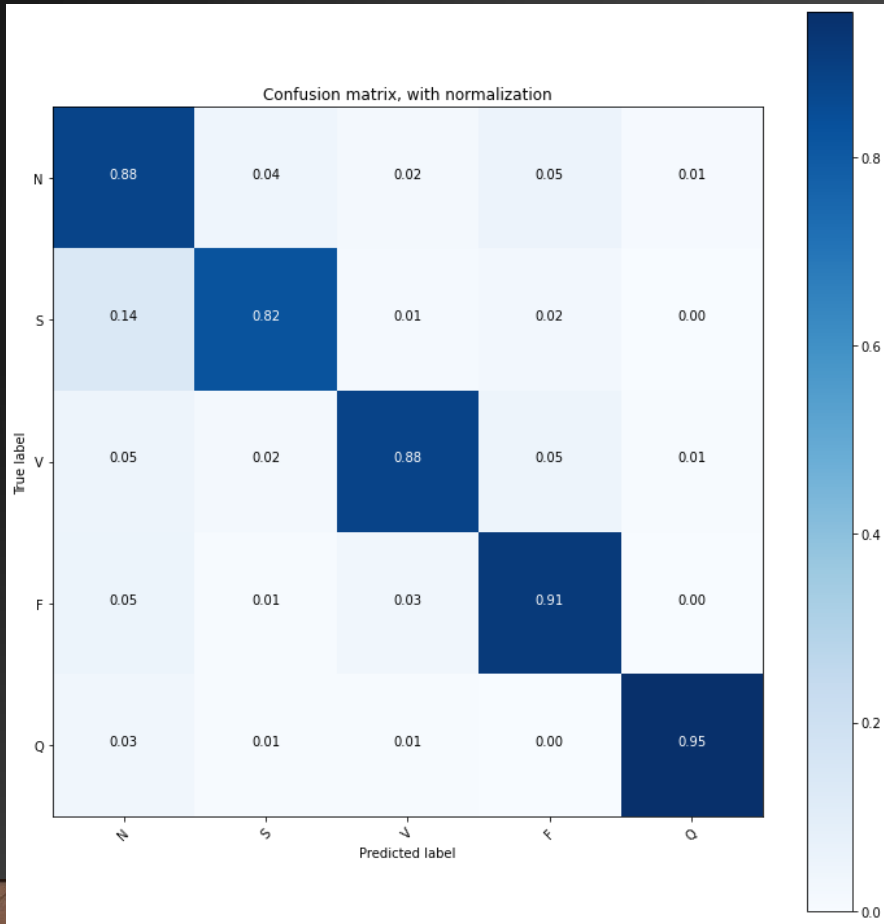
GRADIENT BOOSTING

	N°	Dataset	Manejo de desbalance	Recall c/clases					Peor coeficiente (Matriz confusion)	Obs
				0	1	2	3	4		
GB	1	Completo		0,88	0,1	0,27	0,03	0,69	0,03	
	2	Completo	Undersampling	0,66	0,71	0,77	0,92	0,94	0,66	

REDES NEURONALES (CONVOLUCIONALES)

		Dataset	Manejo de desbalance	Funcion de activacion	Dropout layer	Recall c/clases					Peor coef. Matriz	Observaciones
						0	1	2	3	4		
NL	1	Completo	No	relu	No	1,00	0,12	0,57	0,00	0,82	0,00	
	2	Completo	SMOTE	relu	No	0,83	0,54	0,54	0,81	0,87	0,54	
	3	.iloc[:,0:140]	SMOTE	sigmoid	No	0,65	0,68	0,77	0,94	0,90	0,68	
	4	.iloc[:,0:140]	SMOTE	relu	No	0,69	0,72	0,75	0,96	0,92	0,69	
	5	.iloc[:,0:140]	SMOTE	relu	Si	0,63	0,78	0,68	0,88	0,93	0,68	Gráfica Loss indicio overfitting
	6	.iloc[:,0:140]	Undersampling	relu	Si	0,78	0,81	0,84	0,92	0,97	0,78	Gráfica Loss indicio suavizada

MODELO SELECCIONADO: RANDOM FOREST UNDERSAMPLING RANDOMSEARCHCV



	precision	recall	f1-score	support
0.0	0.99	0.88	0.93	18118
1.0	0.37	0.82	0.51	556
2.0	0.77	0.88	0.82	1448
3.0	0.13	0.91	0.22	162
4.0	0.90	0.95	0.93	1608
accuracy			0.88	21892
macro avg	0.63	0.89	0.68	21892
weighted avg	0.94	0.88	0.91	21892

<https://ml-proyecto-final-ecg.herokuapp.com/>

REFERENCIAS:

- <https://www.kaggle.com/code/gregoiredc/arrhythmia-on-ecg-classification-using-cnn>
- <https://www.kaggle.com/code/abhigyandatta/ecg-heartbeat-classification-smote/notebook>
- <https://www.kaggle.com/code/gregoiredc/arrhythmia-on-ecg-classification-using-cnn/notebook>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=random+forest>

REFERENCIAS:

- <https://www.my-ekg.com/generalidades-ekg/papel-ekg.html>
- <https://www.aprendemachinelearning.com/clasificacion-con-datos-desbalanceados/>
- <https://www.kaggle.com/code/residentmario/under-sampling-and-oversampling-imbalanced-data/notebook>