

01

Introducción

02

Análisis exploratorio

03

Atributos propuestos

04

Modelos utilizados

4.1

Descripción del protocolo experimental.

4.2

Estimación de parámetros.

05

Resultados

06

Conclusiones



Introducción

HT_Sensor_metadata.dat

id	date	class	t0	dt
0	07-04-15	banana	13.49	1.64
1	07-05-15	wine	19.61	0.54
2	07-06-15	wine	19.99	0.66
3	07-09-15	banana	6.49	0.72
4	07-09-15	wine	20.07	0.53
...

HT_Sensor_dataset.dat

id	time	R1	RX	R8	Temp.	Humidity
0	-0.999750	12.862100	...	8.739010	26.225700	59.052800
0	-0.999472	12.861700	...	8.739080	26.230800	59.029900
0	-0.999194	12.860700	...	8.739150	26.236500	59.009300
0	-0.998916	12.860200	...	8.739360	26.241600	58.990500
0	-0.998627	12.859500	...	8.739860	26.246200	58.973600
...		

02

Análisis exploratorio

Gráfico de barras para mostrar la distribución de las clases.

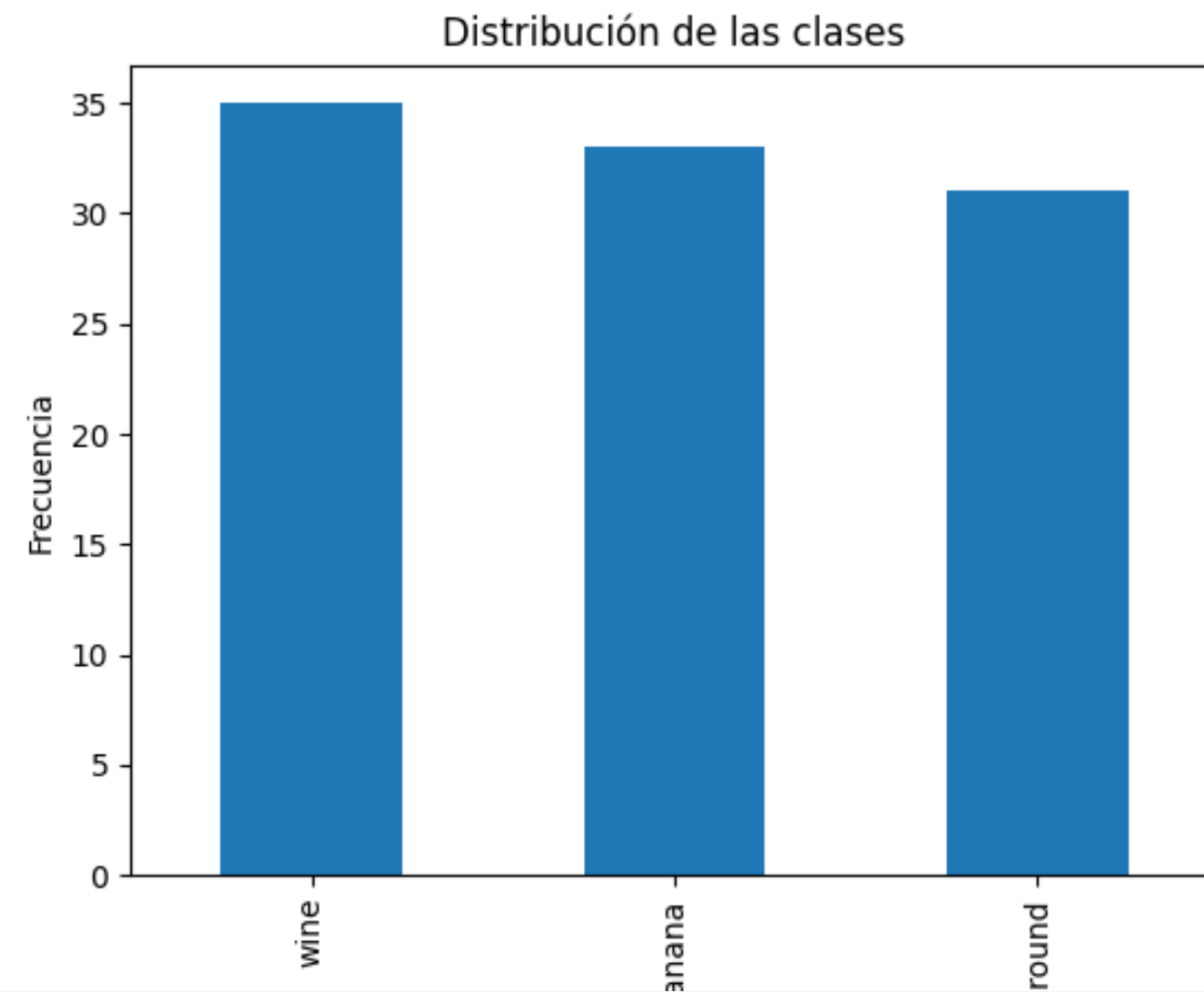


Gráfico de barras para mostrar la distribución de las clases.

02

Análisis exploratorio

Gráficos

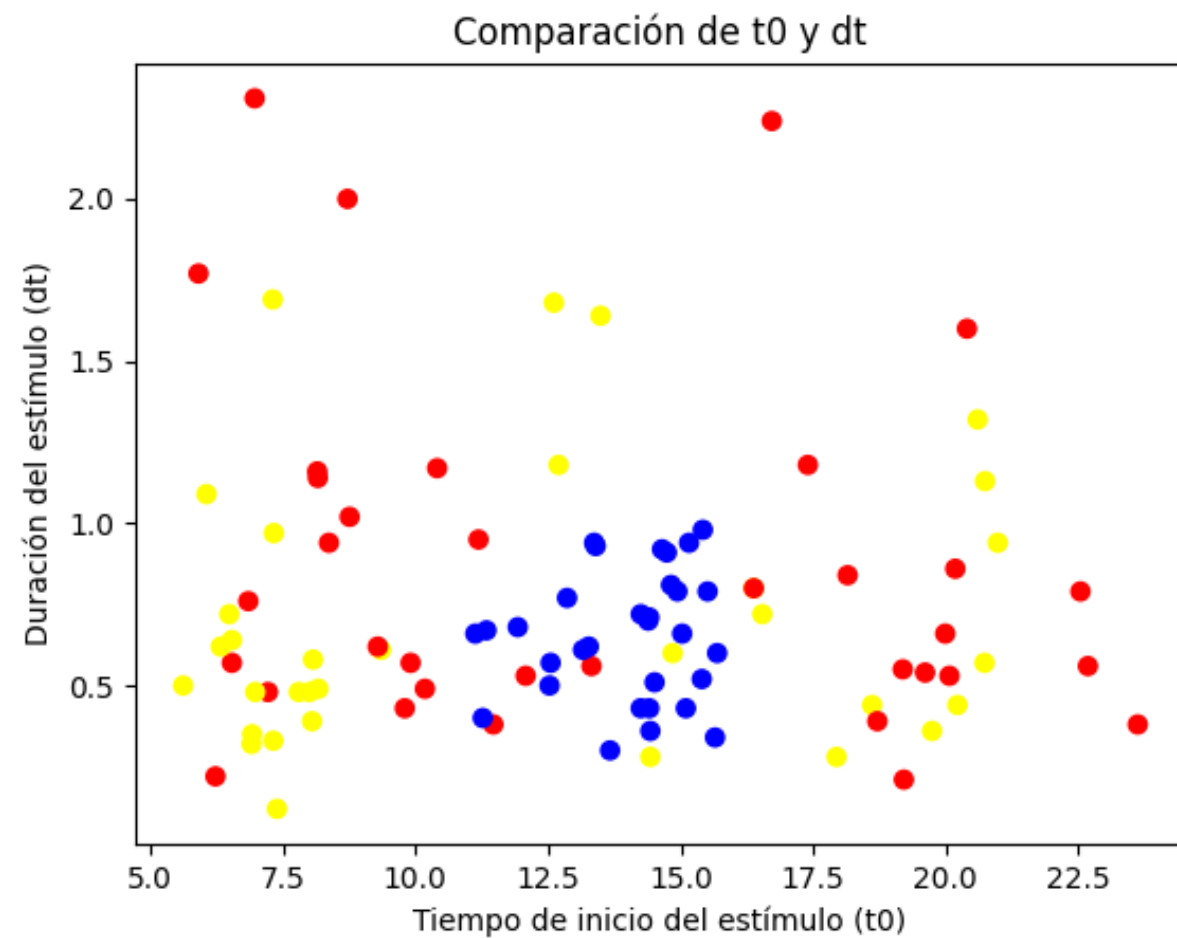


Gráfico de dispersión para comparar la relación entre el tiempo de inicio del estímulo (t_0) y la duración del estímulo (dt)

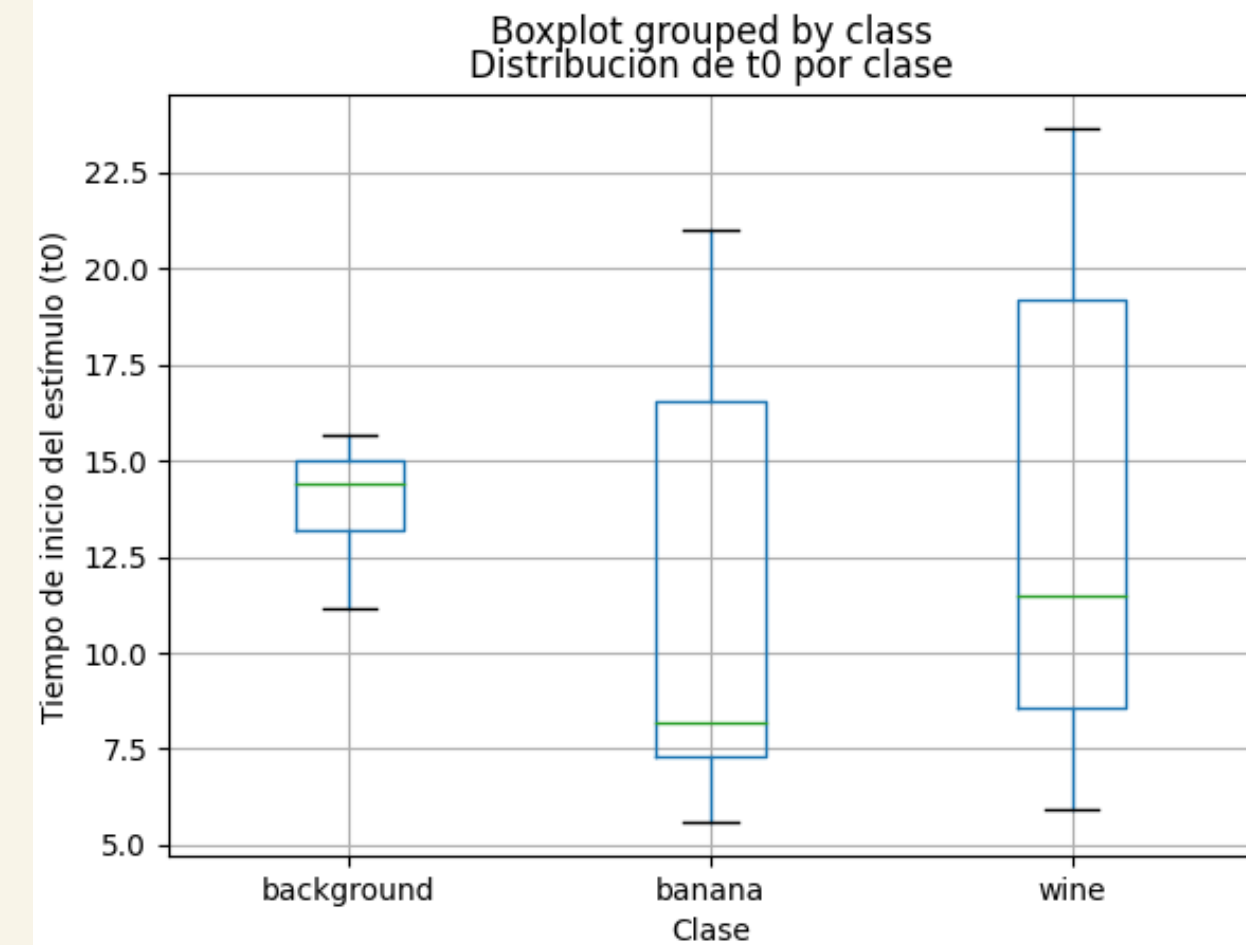


Gráfico de caja para mostrar la distribución del tiempo de inicio del estímulo (t_0) para cada clase

03

Atributos propuestos

Gráficos

Comparación de sensores vs tiempo en banana

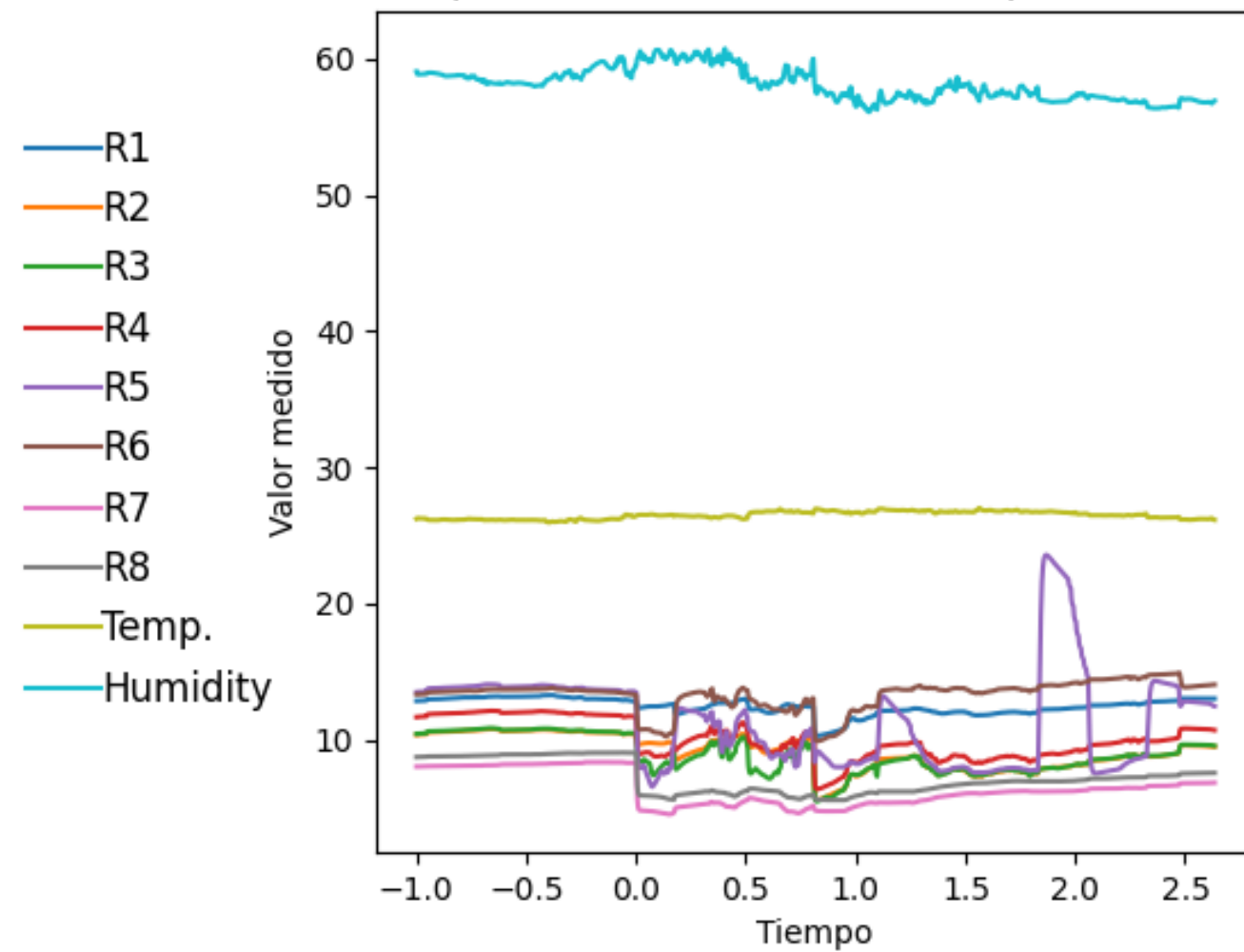


Gráfico que muestra el valor medido por los sensores para cada instante de tiempo. El tipo de sustancia es "banana".

Comparación de sensores vs tiempo en banana despues de asignar valores

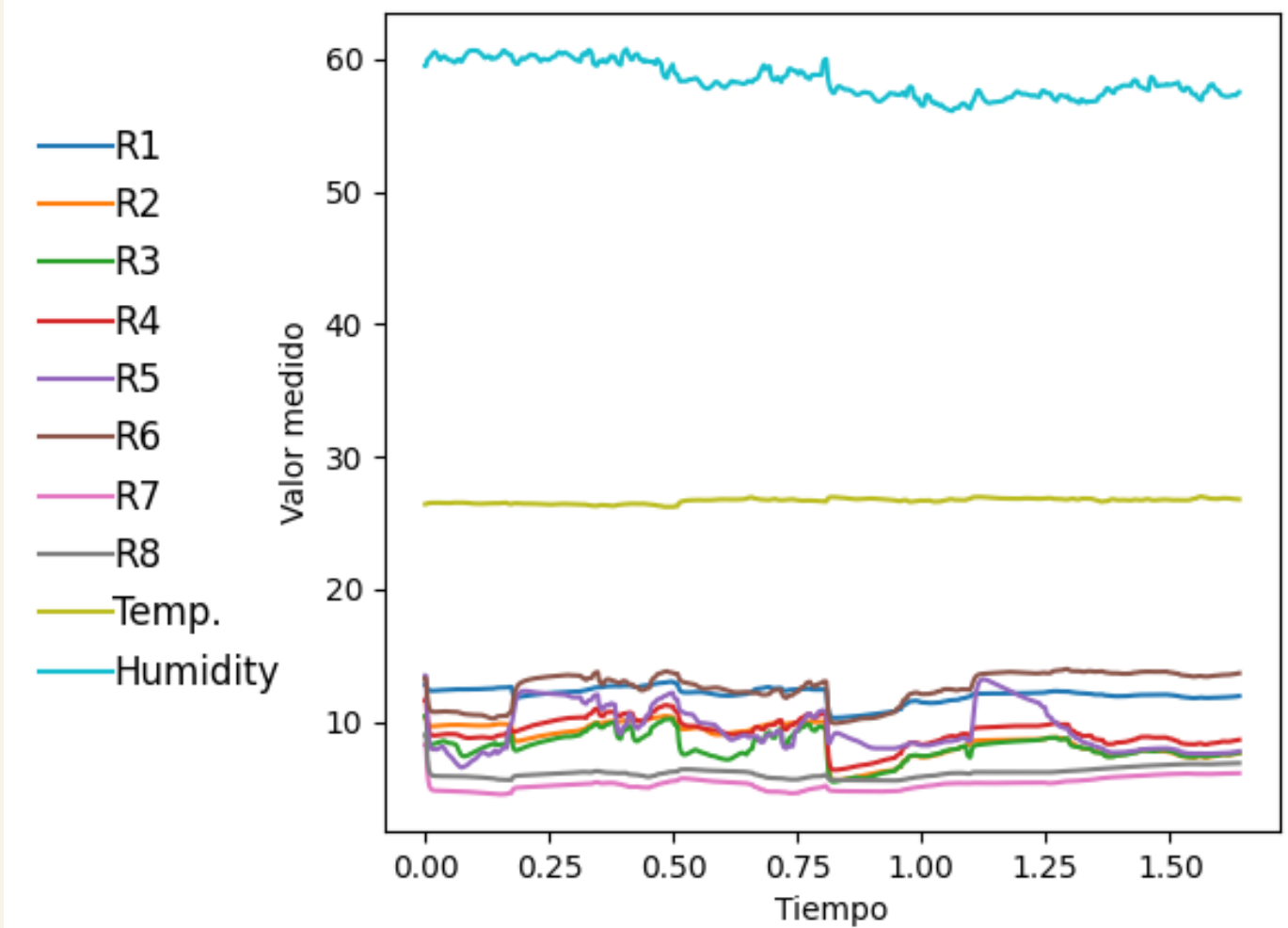


Gráfico que muestra el valor de los sensores para el intervalo en el que se ha determinado que está presente el estímulo.

03

Atributos propuestos

Gráficos

Comparación de sensores vs tiempo en vino

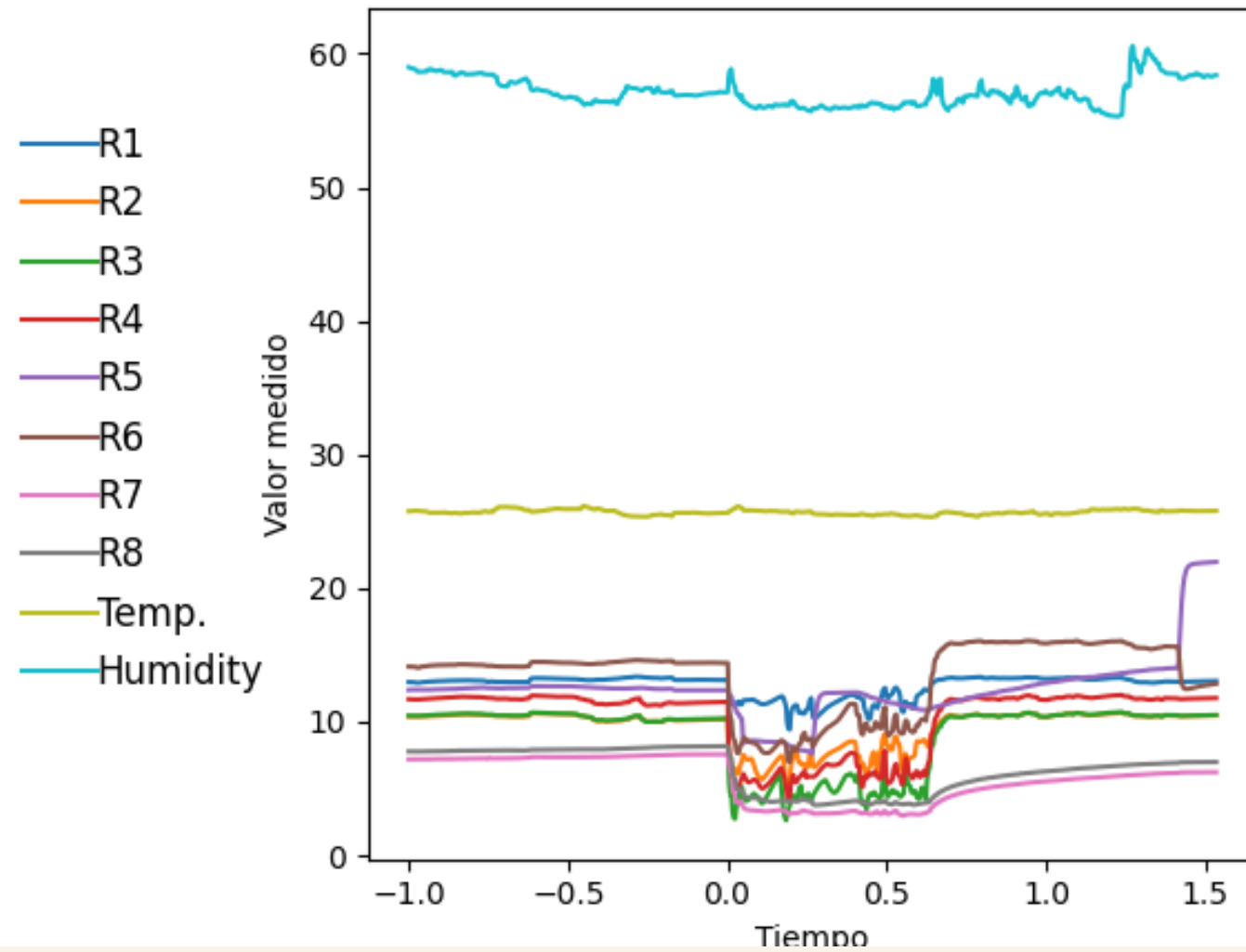


Gráfico que muestra el valor medido por los sensores para cada instante de tiempo. El tipo de sustancia es "vino".

Comparación de sensores vs tiempo en vino despues de asignar valores

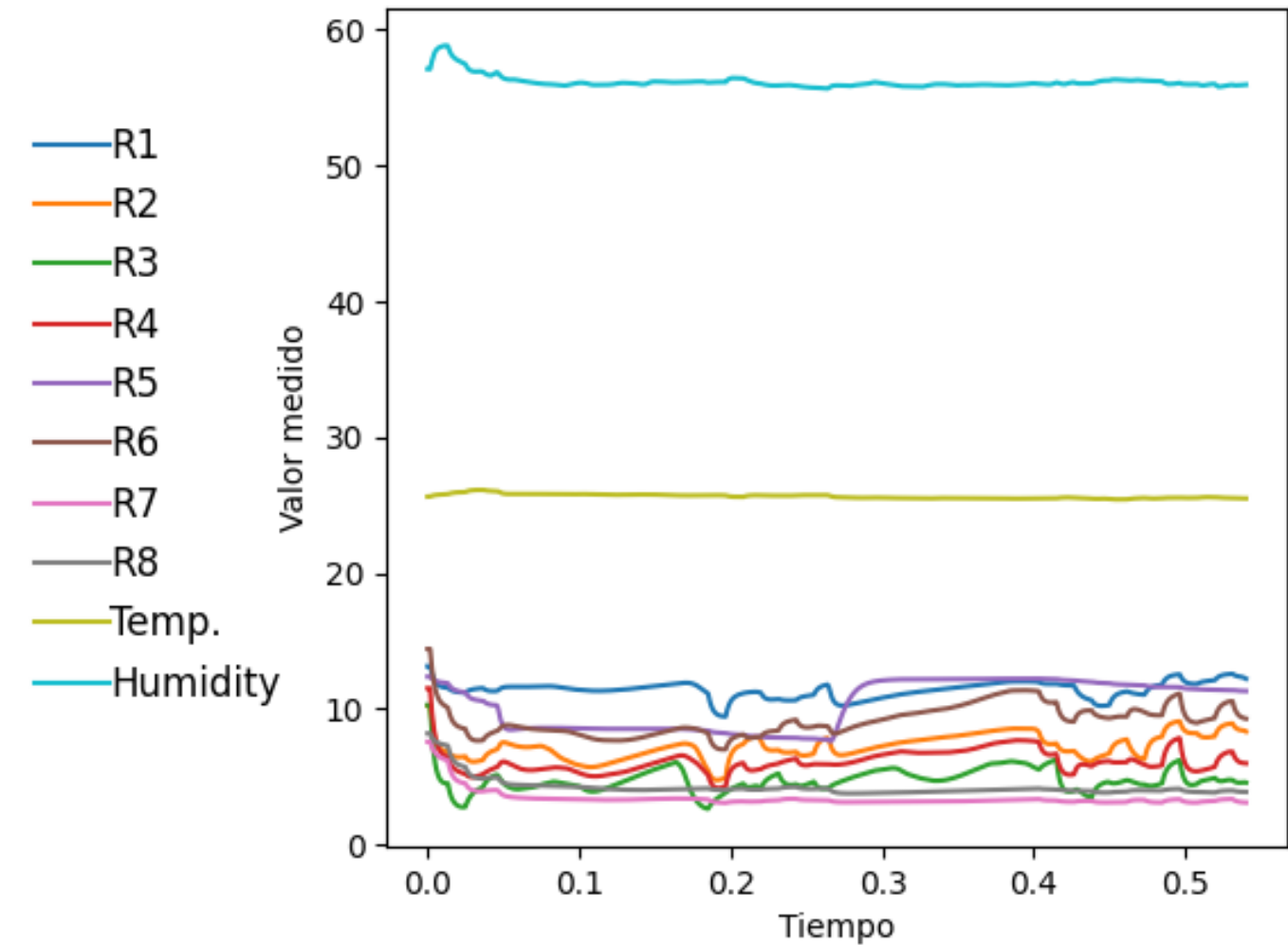
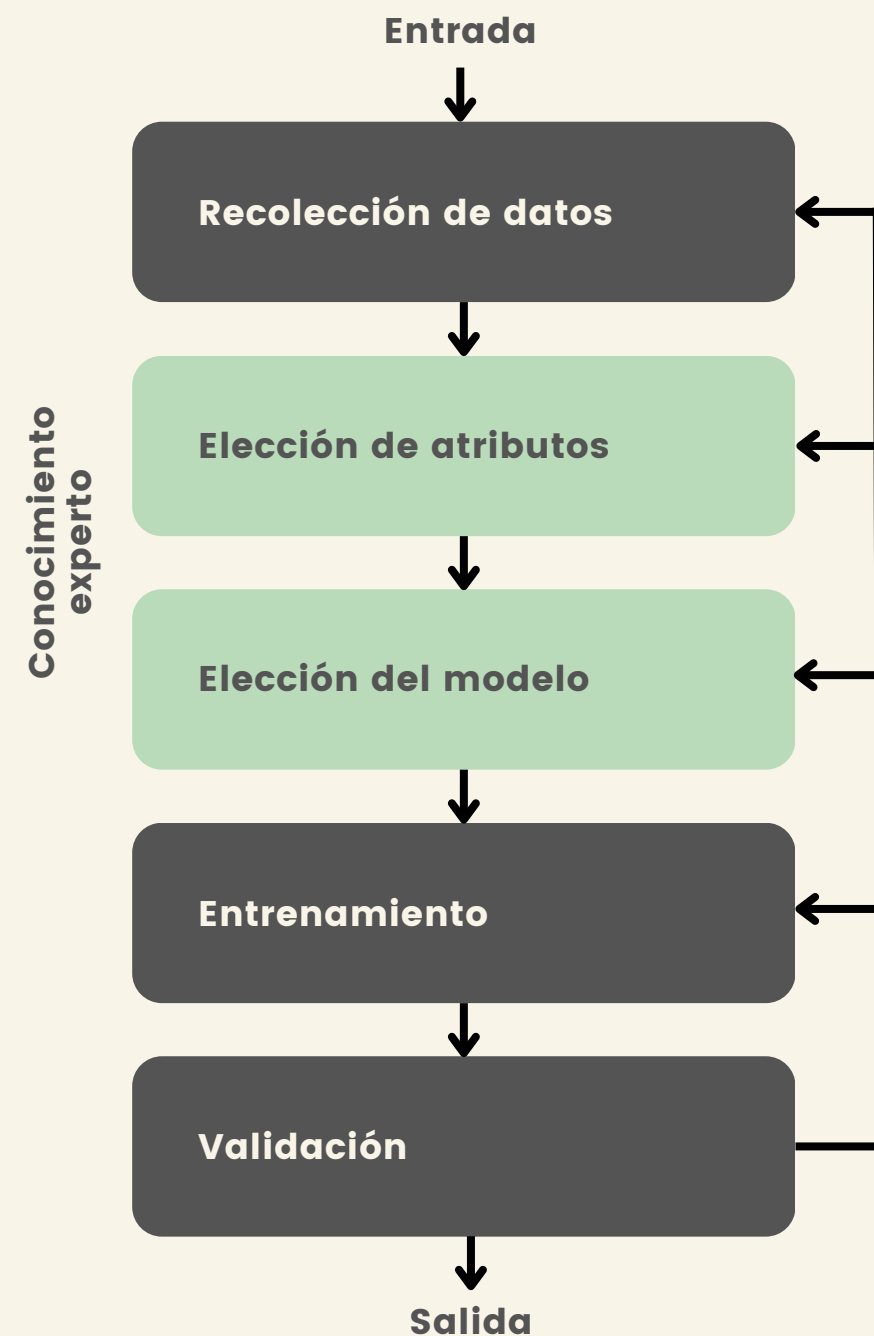


Gráfico que muestra el valor de los sensores para el intervalo en el que se ha determinado que está presente el estímulo.

04

Modelos utilizados

4.1 Descripción del protocolo experimental.



Metodología

- 1) Elegimos atributos para cada experimento por los valores de sus sensores (máximos, mínimos, medios), añadiendo o eliminando columnas si es necesario.
- 2) Ejecutamos tests donde aplicamos distintos modelos a los datasets, examinando y representando los porcentajes de error.
- 3) Volvemos al primer paso, aplicando lo aprendido para realizar cambios que minimicen el error de ejecuciones futuras.

04

Modelos utilizados

4.1 Descripción del protocolo experimental.

Scripts y código fuente

dataset_estimulo.py

Genera un dataset cuya clase toma el valor 1 si el estímulo está presente y 0 si no está presente para cada instante de tiempo.

dataset_clasificar.py

Genera un dataset cuya clase toma el valor del tipo de sustancia detectada. Incluye atributos que identifican cada experimento por el valor de sus sensores. (medios, máximos y mínimos...)

Makefile

Automatiza el proceso de descargar el dataset inicial, procesarlo, realizar tests de modelos y limpiar restos.

test_clasificadores.py

Particiona los datos en entrenamiento y test y aplica diferentes modelos a ambos datasets, mostrando por pantalla los porcentajes de error.

Otros scripts

Utilizamos scripts para descargar el dataset inicial, convertirlo a CSV, generar gráficos y otras funcionalidades.

04

Modelos utilizados

4.2 Estimación de parámetros.

Pruebas:

Para elegir los mejores atributos se han usado la media del porcentaje de acierto en los test.

Clasificador/Atributos	Máximo y mínimo	Máximo, mínimo, desviación típica, media y mediana	Desviación típica, media y mediana
DecisionTreeClassifier	82.98%	79.47%	69.95%
KNeighborsClassifier	60.56%	66.74%	65.62%
RandomForestClassifier	81.20%	79.29%	75.77%
GaussianNB	80.70%	75.98%	74.37%
LogisticRegression	80.13%	77.89%	78.80%

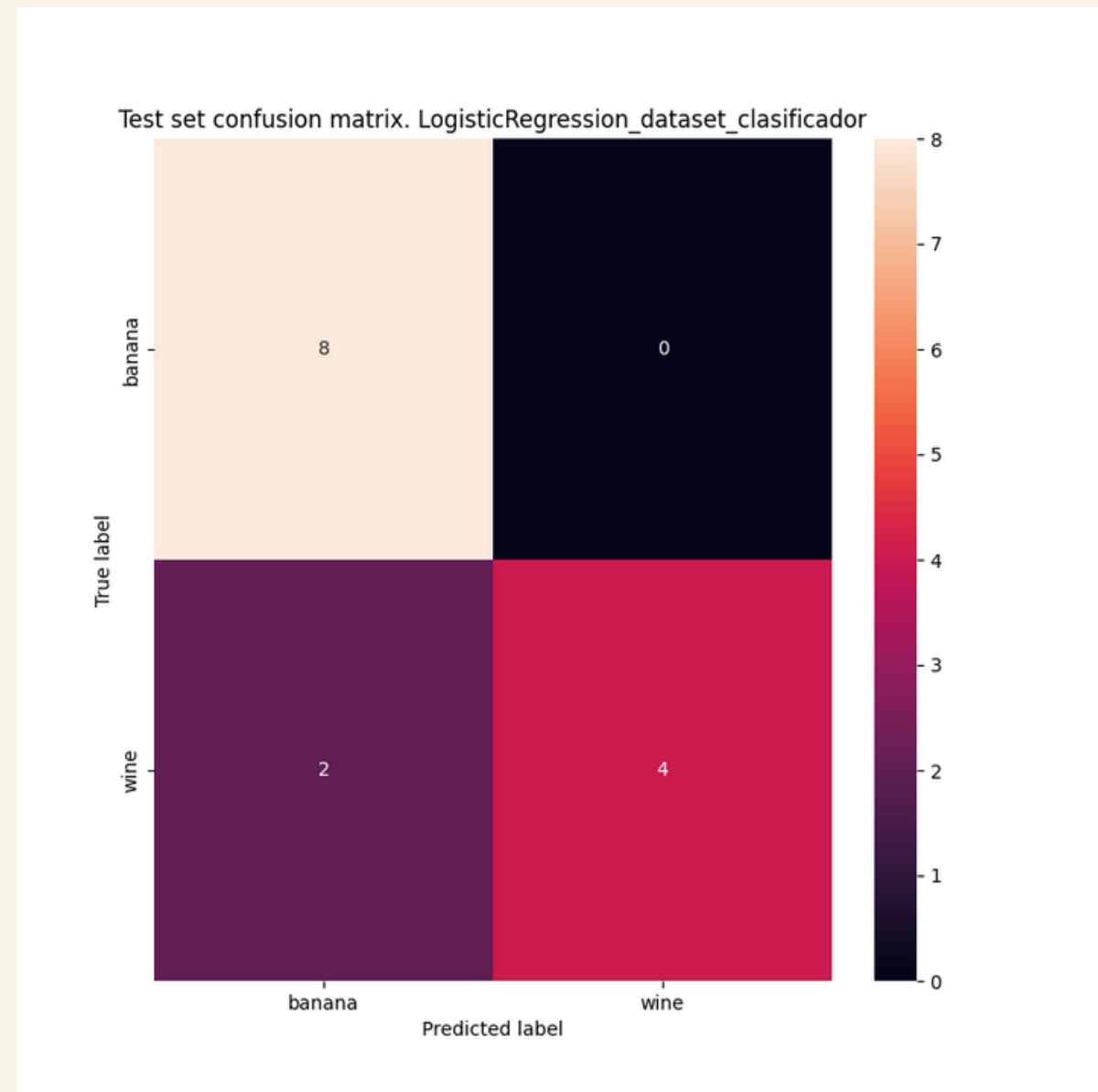
Selección de información útil:

Se hicieron test eliminando la información de distintos sensores, la temperatura y la humedad. Así se pudo comprobar que elementos añadían ruido a los datos.

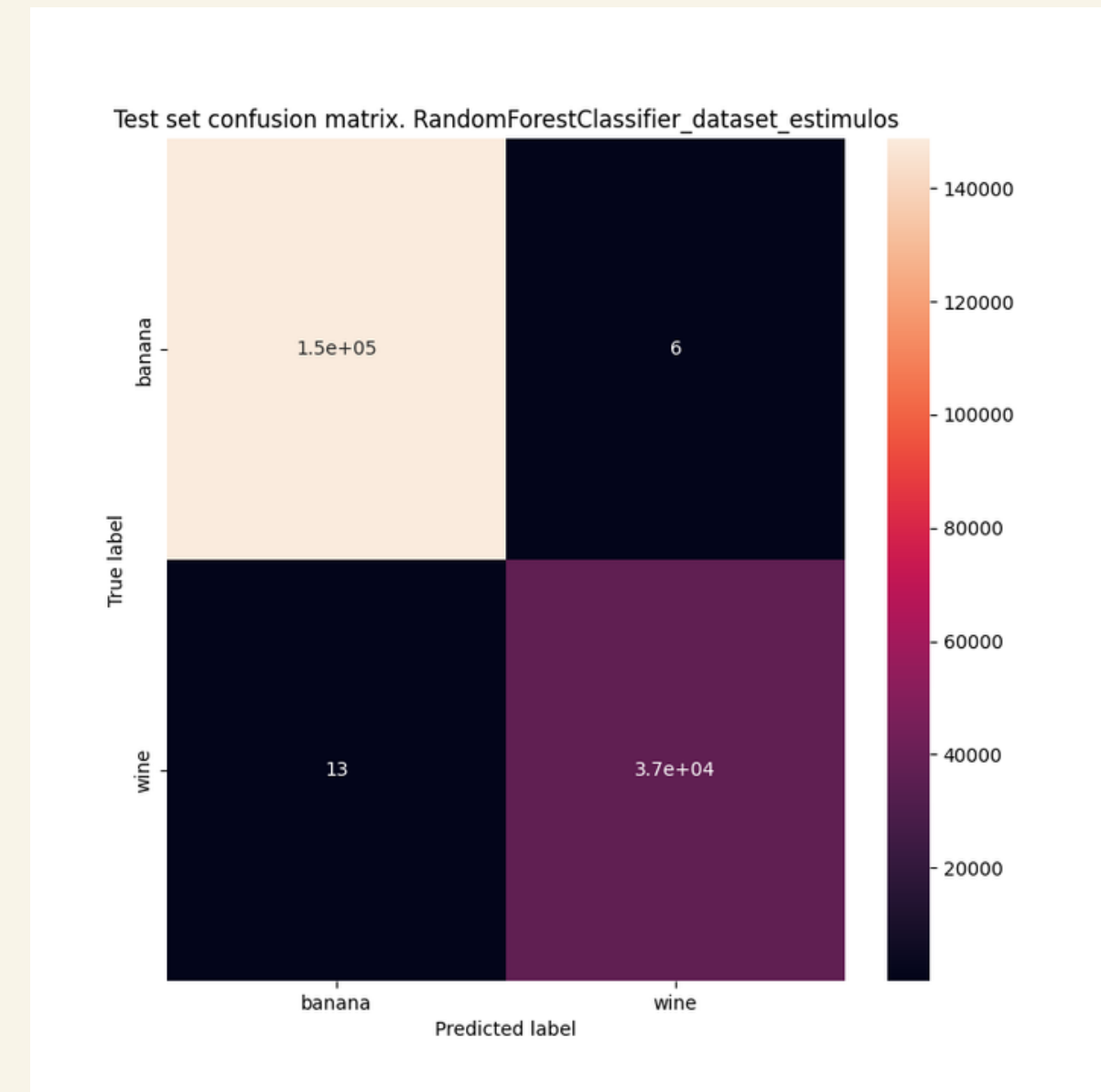
Mejor elección:	Máximo y mínimo sin: sensor 1, sensor 5, humedad ni temperatura
DecisionTreeClassifier	80.56%
KNeighborsClassifier	74.79%
RandomForestClassifier	82.44%
GaussianNB	82.31%
LogisticRegression	83.22%

05

Resultados



Matriz de confusión del modelo Regresión Logística aplicado para clasificar el tipo de sustancia.



Matriz de confusión del modelo Random Forest aplicado para detectar la presencia del estímulo.

1) Se usaron dos modelos diferentes, Random Forest y Regresión Logística, para resolver dos tipos distintos de problemas, clasificar sustancias y detectar presencia de estímulo.

2) Los atributos que definen mejor un experimento son los valores máximos de los sensores, eliminando la humedad y la temperatura, así como los sensores 1 y 5.

3) Los resultados obtenidos son buenos en comparación con los presentados por otros estudios, pero sería interesante aplicar técnicas de filtrado y otros modelos para intentar mejorarlos.