

Monitorización de actividad de gases

UAM - Proyecto de Fund. Aprendizaje Automático - 2020/2021

David Cabornero Pascual

david.cabornero@estudiante.uam.es

José Manuel Chacón Aguilera

josem.chacon@estudiante.uam.es

Mario García Pascual

mario.garciapascual@estudiante.uam.es

Alejandro Santorum Varela

alejandro.santorum@estudiante.uam.es

Contenido

- Definición del proyecto
- Análisis exploratorio de los datos
- Preprocesamiento de los datos
- Datasets utilizados
- Métodos de clasificación
- Resultados obtenidos
- Conclusiones

Definición del proyecto



- Desarrollo constante **iterando** las **fases** habituales de un **proyecto** de aprendizaje automático.
- **Objetivo:** modelar la **monitorización de gases a tiempo real** a través de varios sensores

- Uso de varios **modelos** de la librería de **Scikit-learn**



Análisis exploratorio de los datos

	R1	R2	R3	R4	R5	R6	R7	R8	Temp.	Hum.
vi	11.27	7.12	6.52	8.08	13.64	14.65	3.67	4.13	27.09	57.10
pl	12.13	9.09	8.82	10.16	19.18	15.06	5.28	6.42	27.07	57.75
bg	12.32	9.21	9.31	10.42	14.95	16.36	5.65	6.12	27.33	57.62

- **100 series temporales** con mediciones de **10 sensores**.
- 3 clases a distinguir: **vino, plátano y background**.
- **Desbalance** de ejemplos de **cada clase**.

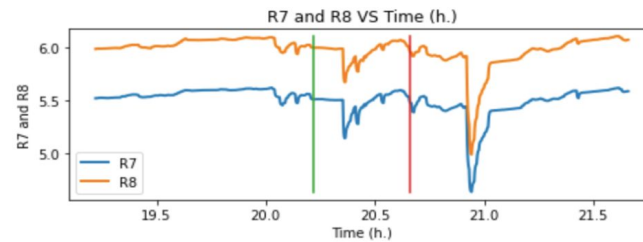
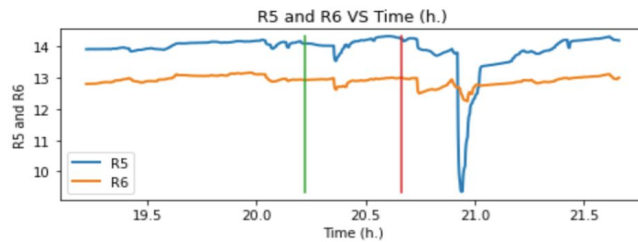
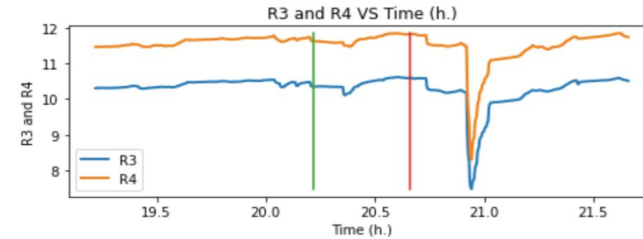
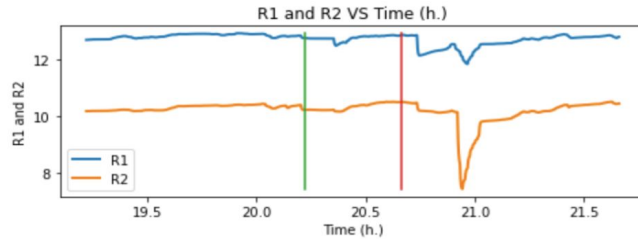
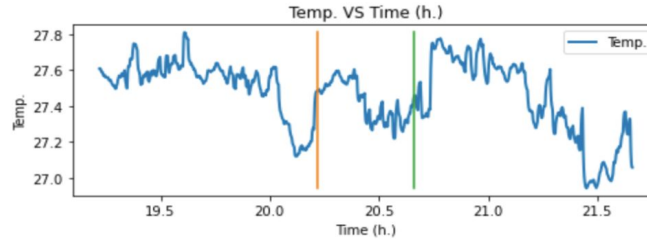
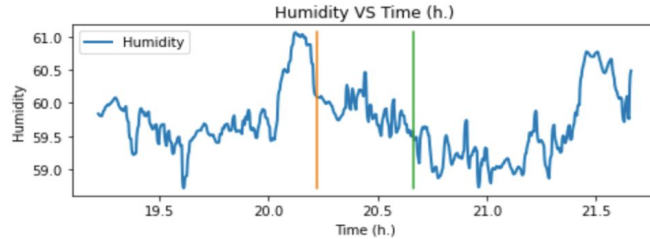
Background → 79.99 %

Vino → 11.57 %

Plátano → 8.44 %

El gran reto

Sensor Reading on Day 18 (banana)



Tratamiento de datos

- **Corrección inicio y final** de los estímulos.
- **Estandarización** los datos si es necesario.
- **Síntesis** de datos nuevos de las **clases minoritarias**.

- **No mezclar** lecturas de la **misma serie** en el **train y test set**.
- La **precisión no es suficiente** para validar con clases desbalanceadas.
 - Uso adicional de **F1-Score**.
- **Evitar** fuerte **dependencia** de la **probabilidad a priori**.

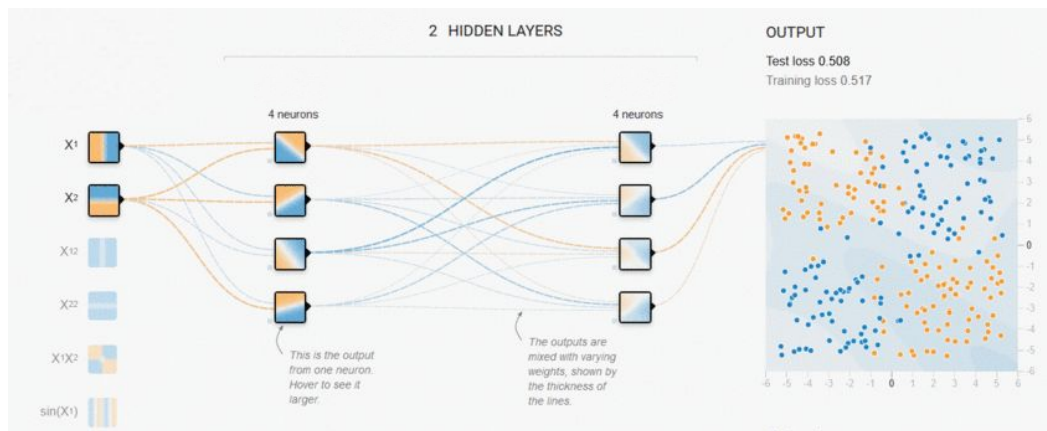
Datasets utilizados

- *Dataset* **original**
- *Dataset* con **estímulos corregidos**
- *Dataset* de **ventanas** móviles (~2 mins)
- *Dataset* con **secuencias** de muestras (~2 mins)
- Uso de **SMOTE**: síntesis de **clases minoritarias**



Modelos de clasificación

- Redes **neuronales** y **Bagging**
- Métodos **Kernel** y **Bagging**
- Conjuntos de árboles de decisión:
 - *Random Forest*
 - *Boosting*
- Redes **convolucionales**
- Redes **recurrentes**



Resultados obtenidos

- Mejora usando **ventanas**.
- Mejor rendimiento con **ensembles**.
- La **élite** se sitúa en torno al **86-88%**.

MODELO	Val. acc.	Val. F1-Sre.	Test acc.	Test F1-Sre.
RNN	83.1% \pm 4%	78.6% \pm 6%	81.7%	77.0%
CNN-1D	81.8% \pm 5%	76.7% \pm 7%	84.0%	79.4%
CNN-2D	81.9% \pm 3%	78.4% \pm 3%	83.0%	79.3%

Precisión

MODELO	CL-DB	CL-DB-Sm	MW-DB	MW-DB-Sm
NN	84.6% \pm 3%	- - -	84.9% \pm 2%	82.2% \pm 3%
Ens. NN	84.4% \pm 6%	- - -	85.7% \pm 3%	86.2% \pm 3%
Kernels	79.1% \pm 2%	- - -	82.8% \pm 3%	82.6% \pm 3%
Ens. Kernels	80.8% \pm 3%	- - -	85.6% \pm 3%	84.8% \pm 4%
Rand. Forest	84.1% \pm 5%	83.1% \pm 5%	87.8% \pm 4%	85.6% \pm 3%
AdaBoost	83.3% \pm 1%	83.7% \pm 1%	86.3% \pm 1%	87.6% \pm 2%

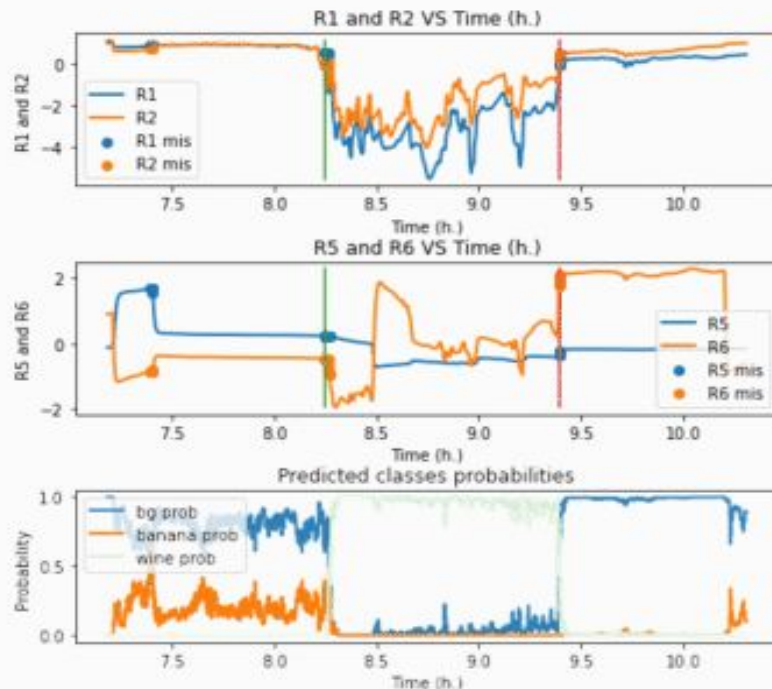
F1-Score

MODELO	CL-DB	CL-DB-Sm	MW-DB	MW-DB-Sm
NN	83.5% \pm 3%	- - -	84.2% \pm 2%	83.0% \pm 3%
Ens. NN	83.0% \pm 7%	- - -	84.2% \pm 3%	86.3% \pm 3%
Kernels	73.4% \pm 3%	- - -	79.5% \pm 4%	82.2% \pm 3%
Ens. Kernels	74.9% \pm 2%	- - -	82.7% \pm 3%	84.9% \pm 3%
Rand. Forest	80.3% \pm 5%	81.6% \pm 5%	84.3% \pm 3%	85.1% \pm 4%
AdaBoost	79.7% \pm 2%	81.4% \pm 2%	83.6% \pm 1%	84.3% \pm 2%

Resultados obtenidos II

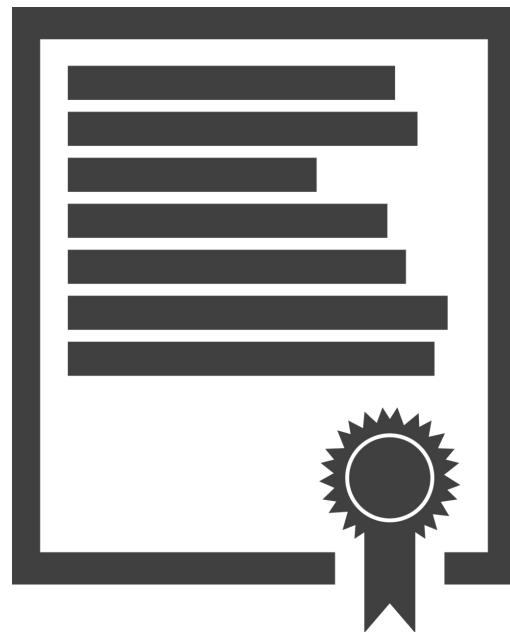
- Con el uso de **SMOTE** el **vino** es claramente **diferenciable**.
- **SMOTE** potencia el rendimiento en el estímulo *banana*.
- **Bagging** de redes neuronales aportan los **mejores resultados**.

RANDOM FOREST SMOTE
Sensor reading on day 29 (Wine)



Conclusiones

- Aumento del ***F1-Score*** con **ventanas** y **SMOTE**.
- **Mejores clasificadores:**
 - Random Forest
 - Bagging con Redes Neuronales
- **Siguientes** pasos:
 - Nueva **recolección de atributos**
 - **Pulir modelos** más complejos



¡Gracias por su atención!



David Cabornero Pascual

david.cabornero@estudiante.uam.es

Mario García Pascual

mario.garciapascual@estudiante.uam.es

Gonzalo Martínez Muñoz

gonzalo.martinez@uam.es

José Manuel Chacón Aguilera

josem.chacon@estudiante.uam.es

Alejandro Santorum Varela

alejandro.santorum@estudiante.uam.es