

APRENDIZAJE AUTOMATICO:

ACOUSTIC EXTINGUISHER FIRE DATASET

INTRODUCCIÓN

Acoustic Extinguisher Fire Dataset es una base de datos que presenta una idea muy interesante.

¿Y si se pudiese apagar un fuego usando tan solo sonido?

Esta surge en contraposición a otros métodos como el agua, muy ineficiente contra fuegos alimentados por óleos, o espuma de extintores, ambos materiales no reutilizables una vez utilizados.

Intrigados por esta idea, hemos decidido poner a prueba distintos modelos de aprendizaje automático para determinar cuan posible es hacer predicciones consistentes sobre la efectividad y/o seguridad de la aplicación de este método de regulación de llamas.

AUTORES

Igor Vons
Aitor Uranga

FUENTES

KAGGLE:Acoustic Extinguisher Fire Dataset
KOKLU M., TASPINAR Y.S., (2021). Determining the Extinguishing Status of Fuel Flames With Sound Wave by Machine Learning Methods. IEEE Access, 9, pp.86207–86216, Doi: 10.1109/ACCESS.2021.3088612
TASPINAR Y.S., KOKLU M., ALTIN M., (2021). Classification of Flame Extinction Based on Acoustic Oscillations using Artificial Intelligence Methods. Case Studies in Thermal Engineering, 28, 101561, Doi: 10.1016/j.csite.2021.101561
TASPINAR Y.S., KOKLU M., ALTIN M., (2022). Acoustic-Driven Airflow Flame Extinguishing System Design and Analysis of Capabilities of Low Frequency in Different Fuels. Fire Technology, Doi: 10.1007/s10694-021-01208-9

OBJETIVO

Evaluar tanto la validez del dataset para realizar predicciones precisas como la aplicabilidad de distintas técnicas de aprendizaje automático estudiadas y desarrolladas a lo largo del curso. En particular se hablará de los siguientes métodos:

- Regresión Logística
- Redes Neuronales
- Naive Bayes
- Support Vector Machines (SVM)
- Ensembles (en particular DecisionTree, Bagging, Boosting y Random Forest junto con OVO y OVA)

METODOLOGÍA

Para poder evaluar los clasificadores normalmente haremos uso de hiperparámetros que definan su comportamiento. Hacer una búsqueda manual de los mejores para un caso particular es muy laborioso, por ello antes de poner a prueba cada uno de ellos realizaremos un *grid search*, una técnica que permite obtener la mejor combinación entre los posibles valores de cada parámetro (al menos los que suministremos). Una vez los hayamos obtenido, podemos proceder a utilizarlos y probar su efectividad. Cabe comentar el no uso de K-means, ya que se ha considerado que debido a la forma del dataset (no parece seguir la estructura de clúster que requiere K-means) y ser la separación en solo dos clases, no aporta ninguna ventaja en especial aquí.

LEYENDA:

- para combustibles líquidos
 - TAMAÑO: 7, 12, 14, 16, 20 cm (Registrado como 7 cm=1, 12 cm=2, 14 cm=3, 16 cm=4, 20 cm=5)
 - COMBUSTIBLE: 1=Gasolina, 2=Queroseno, 3=combustible baja viscosidad
 - DISTANCIA: 10-190 cm
 - DECIBELIOS: 72-113 dB
 - FLUJO DE AIRE: 0-17 m/s
 - FRECUENCIA: 1-75 Hz
 - ESTADO: ● no extinto, ● extinto
- para LPG
 - TAMAÑO: Configuración de flujo medio, Configuración de flujo completo (Registrado como Configuración flujo medio=6, Configuración flujo completo=7)
 - COMBUSTIBLE: 4=LPG
 - DISTANCIA: 10-190 cm
 - DECIBELIOS: 72-113 dB
 - FLUJO DE AIRE: 0-17 m/s
 - FRECUENCIA: 1-75 Hz
 - ESTADO: ● no extinto, ● extinto

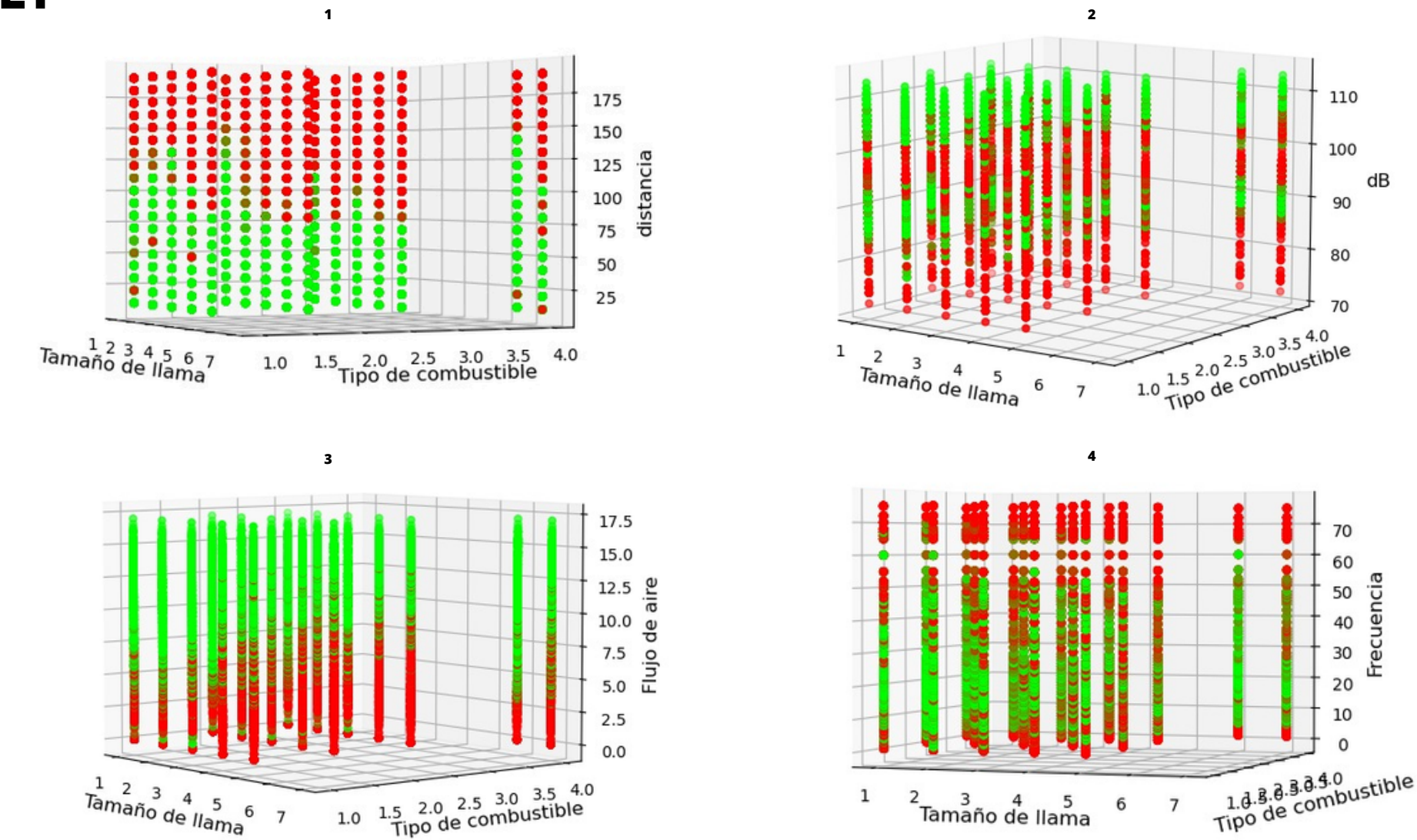
ANÁLISIS VISUAL DEL DATASET

En las gráficas se han representado los datos que conforman el dataset. Los ejes del plano vertical siempre son Tamaño de la llama y Tipo de combustible. El eje vertical en cada una de ellas contrapone una de las 4 variables restantes (Distancia, dB, Flujo de aire y Frecuencia). La distinta inclinación de cada una de ellas se debe a se ha buscado el ángulo desde el que se pueda ver mejor los patrones de los valores.

Se puede observar que en las gráficas 1, 2 y 3 la variable de principal peso es siempre la vertical. Se pueden encontrar "fronteras" de cambio. Distancia: ~>75-100
dB: ~90 y ~110
Flujo de aire: ~>7.5-10

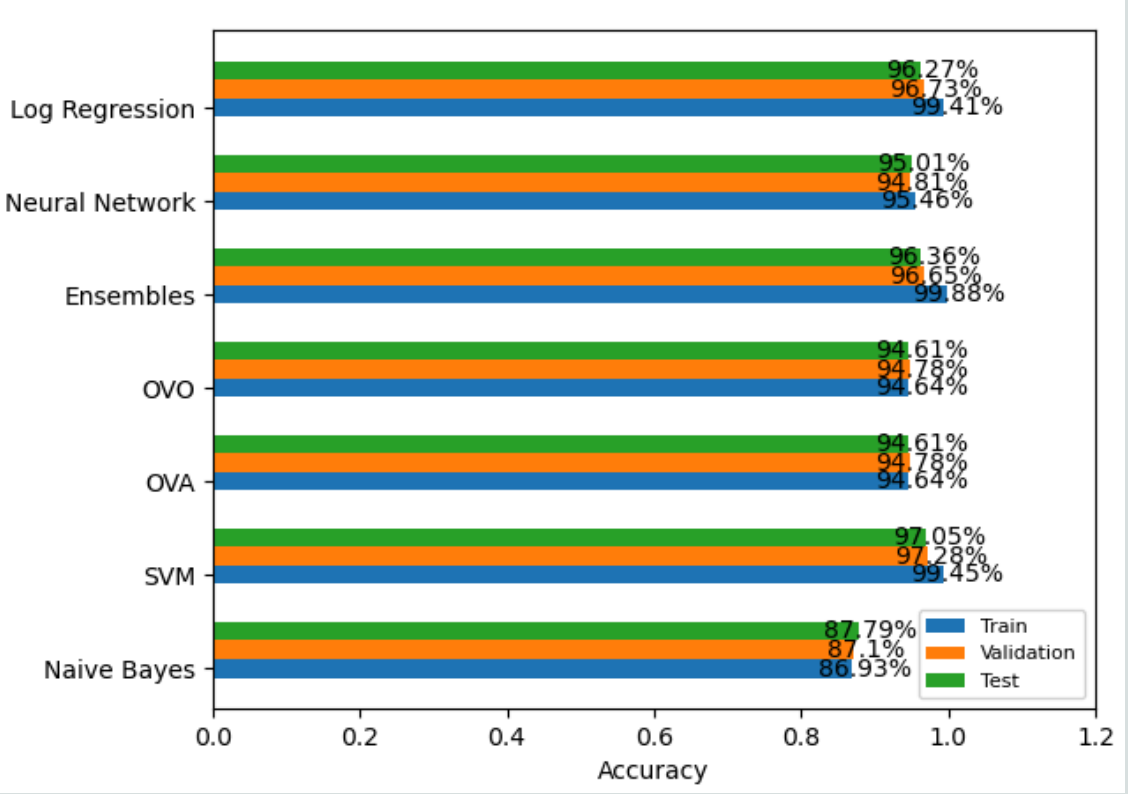
También se aprecia que a menor tamaño de las llamas es más fácil que esta se apague, reflejándose esto en mayor segmento con extinciones exitosas.

La gráfica 4 resalta porque es la única en la que el eje vertical tiene casi menos importancia que los otros dos. Observamos cierta correlación entre frecuencia y resultado, pero notamos mucho más la influencia del tamaño de la llama.



RESULTADOS Y HALLAZGOS

De las gráficas se extrae que las variables de principal importancia son Distancia, dB y Flujo de aire. Esto también nos deja claro que el dataset posee cierta estructura consistente con la que se puede trabajar para extraer predicciones. Esto se refleja también en la gráfica de precisiones, donde hemos podido alcanzar más de un 99% con los ejemplos de entrenamiento o más de 97% para los de evaluación. Los modelos con los que se ha obtenido los mejores resultados son Support Vector Machines, Ensembles (concretamente Bagging con clasificador base siendo un árbol de decisión) y Regresión Logística. El peor resultado con diferencia nos los da Naive Bayes no alcanzando siquiera el 90%, siendo además el único que obtuvo una mejor precisión en test que en train.



CONCLUSIÓN:

Basándonos en los datos obtenidos, el modelo de Support Vector Machines es el más preciso. De todas formas, dependiendo cuan preciso requerimos que sea el modelo, esto puede ser insuficiente. Un 3% de error es inaceptable en el control de un motor de combustión, pero no lo es tanto para un extintor de velas de uso personal.

LÍNEAS A FUTURO:

- Optimización de código (velocidad de ejecución)
- Pruebas de modelos más complejos
- Estudio comparativo de las clasificaciones erróneas