Guión:

Portada:

En este trabajo vamos a hablar de la familia de enfermedades cardiovasculares que en adelante llamaremos ECV’s

P2:

No hace falta que lo leais todo, tan solo es para mostrar la cantidad de enfermedades que se engloban en esta familia.

P3:

Las enfermedades cardiovasculares son la [principal causa de muerte](https://en.wikipedia.org/wiki/List_of_causes_of_death_by_rate) en todo el mundo excepto en África. En conjunto, las ECV provocaron 17,9 millones de muertes (30% aprox) en 2015.

Se estima que hasta el 90% de las enfermedades cardiovasculares pueden prevenirse de una manera fácil como alimentación saludable.

Esto a nivel global, pero también vamos a comprobar que sucede en España con las ECV’s. Para ello he recurrido a información de un reputado Data Scientist español. (broma, compañero de clase).

P4:

Como bien dijo David, este tipo de enfermedades van a la baja en España, pero siguen siendo la principal causa de muerte.

P5:

Y no solo a nivel humano, que ya sería un problema por si solo, a nivel monetario supone perdidas billonarias. las pérdidas económicas acumuladas proyectadas de todas las Enfermedades No Transmisibles son de $7,28 billones. Las ECV’s representan casi el 50% de esta pérdida. Este costo surge de las personas en edad laboral y gente a su alrededor que tiene que cuidarlos.

P6:

¿Qué podemos hacer con esta problemática?

P7:

Pues poner la tecnología al servicio de la sociedad. Machine Learning en este caso.

P8:

El *Machine Learning* es una disciplina del campo de la [Inteligencia Artificial](https://www.iberdrola.com/te-interesa/tecnologia/que-es-inteligencia-artificial) que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones. Vemos soluciones de este tipo en recomendaciones de redes sociales, programas para dar creditos de grandes bancos, en cualquier sitio.

P9:

Mi objetivo con este trabajo es comparar diversos algoritmos de ML hasta encontrar uno que sirva como sistema de apoyo médico en la prevención de ECV’s.

Estos modelos constan de datos simples como edad, sexo, tipo de dolor de pecho y una prueba de esfuerzo con ejercicio de la que un médico puede extraer valiosa información recogida por objetos simples y baratos como el que veis en la foto.

P10:

Se han probado todos los modelos que ahí veis y a la derecha tenemos los 5 mejores resultados en test. Aquí vemos que el mejor es el modelo SVM con una exactitud del 80%. Pero claro, aquí entramos en un problema de métricas. Por supuesto queremos un modelo lo mas exacto posible, pero dentro del 20% de errores podemos marcar a gente como “no enferma” cuando si lo está, cuando sería preferible al revés, ya que cada vida es importante. Por lo tanto, nos interesa maximizar el recall también.

P11:

Si maximizamos el recall vemos que el modelo SVM sigue siendo el mejor con mas de un 90%, pero perdemos en exactitud que baja hasta un 71%.

P12:

Tras una serie de cambios manuales al modelo intentando conseguir una versión intermedia de SVM de los 2 modelos que hemos visto. Consigo un resultado **por encima del 76% de accuracy y por encima del 90% en recall.**

Es decir, tendriamos cubierto un 90% de enfermos.

Tenemos un recall por encima de la media de los otros 4 modelos maximizados por recall. Y una exactitud tambien por encima de la media de los otros 4 modelos maximizados por exactitud.

P13:

Todo esto muy bien, pero ¿que es el modelo SVM?

P14:

Pues simplemente y llanamente esto.

Tenemos dos grupos que clasificar (enfermos y no enfermos), se ponen en una gráfica y se traza una linea que los separe.

Ahora bien, no todos los problemas son tan fáciles.

P15:

Si mirais a la izquierda podemos ver un problema donde no se puede trazar una linea para clasificarlos. ¿Que hacemos? Pues aumentar las dimensiones y cortar con un hiperplano como podeis ver en el grafico. Despues se aplanan esas dimensiones, quedando un círculo en el ejemplo de la izquierda y una linea con curvas en la derecha.

P16:

Otro ejemplo de como funciona el SVM

P17:

Este modelo nos ha dado que las variables mas importantes son

--Reconocer el tipo de dolor de pecho

--Oldpeak y ST\_slope que estan relacionados con la informacion que debe saber leer el medico de la prueba de esfuerzo.

--El sexo, la edad

--La presion arterial

-- y el máximo de pulsaciones al que llega en la prueba de esfuerzo.

Otras variables que necesitarian de otro tipo de pruebas y un Sistema medico de calidad son irrelevantes para este modelo en concreto.

P18:

Estamos ante un enorme problema que afecta al conjunto de la sociedad.

Este proyecto no pretende por si solo paliar el problema, pero como sistema de apoyo a un médico y gracias a la accesibilidad de sus datos podría suponer una gran ayuda en la prevención de este tipo de enfermedades en todo tipo de paises y zonas.

P19:

Como curiosidad técnica, las herramientas de optimización bayesianas como Optuna son un gran punto a favor en machine learning.