**Proyecto Machine Learning**

Pedro Pino Von Tzebrikow de Villardo.

Memoria

* **Introducción**

Empecé este proyecto con la idea de indagar en algún tema que se corresponda con mis estudios anteriores, la psicología. He estado buscando y probando con multitud de diferentes bases de datos que se correspondían con diferentes afecciones de la salud mental. Al final me decidí por analizar la depresión, la ansiedad y el stress y como se relacionan entre ellas.

Una vez decidida la base de datos empecé por corregir el test psicológico como se indica en su guía, esta base de datos contiene los datos del DASS, un test que en psicología es muy utilizado. El DASS consta de 42 ítems que nos dan, una vez corregido, puntuaciones en ansiedad, stress y depresión. Pero en muchas ocasiones se presenta en versiones reducidas que solo nos dan información de una de las variables.

Por lo tanto, se me ocurrió probar a predecir la depresión en función de la ansiedad y el stress, o incluso de solo una de ellas.

- **Desarrollo**

A partir de este punto empiezo el proyecto en sí, lo primero que me propongo es realizar una limpieza de los datos, ya que la base de datos es muy extensa, cubre muchísima información demográfica de cada individuo, y algún test psicológico más que en este caso no me interesan para mi experimento, aunque no descarto realizar pruebas adicionales más adelante.

Una vez acabada la limpieza de datos me propongo corregir el DASS para obtener las puntuaciones totales en ansiedad, stress y depresión, y me doy cuenta de que las puntuaciones son altísimas. Indago un poco en los datos y me doy cuenta de que la mayor parte de personas provienen de regiones pobres de Asia, donde la situación no es la idónea, eso explica porque las puntuaciones son tan altas. Como es muy complicado encontrar bases de datos con información fiable en psicología, y aún más cuando necesitas que tengan información de varios test realizados por la misma persona, decido seguir con el experimento y usar la mediana de la depresión del conjunto de test como punto de corte, en vez de con el que propone la corrección del DASS, ya que en psicología siempre hay que tener en cuenta el entorno en el que están las personas a las que estamos analizando.

En otro momento, mi intención es repetir este modelo con datos obtenidos de forma autónoma, pasando yo mismo los test para analizar este problema en un entorno que me sea más familiar.

Una vez tomada esta decisión y codificadas mis variables, ya solo me quedaba aplicar el modelo, he utilizado una regresión polinómica y he usado como métrica de scoring el recall, ya que veo este modelo como una lupa a poner sobre personas en riesgo de sufrir depresión, y quiero minimizar los falsos negativos.

Una vez obtenidos mis resultados, paso a sacar mis conclusiones.

* **Librerias Utilizadas.**
* Pandas
* Numpy
* Matplotlib
* Seaborn
* Sklearn
* Warnings
* **Dificultades Encontradas**

Este proyecto me ha hecho darme cuenta de que el Machine Learning, y el hecho de predecir variables me apasiona, pero también que requiere de mucha práctica, que ahora mismo tengo todos los algoritmos algo mezclados y que voy a tener que hacer muchos proyectos como este para sentirme totalmente cómodo realizándolos a la vez que entiendo perfectamente lo que estoy haciendo, porque lo estoy haciendo y los resultados que obtengo.

* **Conclusión**

Los resultados del modelo son muy interesantes, ya que predice de forma bastante acertada la puntuación en depresión a partir de las puntuaciones en las otras variables. Esto puede ser muy interesante para que en algunas empresas o en algunos colegios o institutos en los que solo se pasan test de ansiedad o de stress se pueda poner la lupa encima de ciertas personas y que reciban ayuda sin la necesidad de hacerles pasar por más pruebas o de rellenar más test que sus compañeros.