

Tema: Data Science, Predicción de la calidad del aire.

Objetivo: Utilizar la información brindada para aplicar métodos de análisis y predicción de datos. El participante aprenderá a crear un modelo de predicción básico, utilizará los datos brindados para leerlos y analizarlos y al final podrá proponer una solución a algún problema que haya detectado gracias al análisis que hizo previamente.

Problema: Se les dará un set de datos que indican diferentes parámetros sobre la calidad del aire en una ciudad. Su reto es predecir la contaminación del aire que tendrá la ciudad en un tiempo futuro.

Datos: Los datos se encuentran en la misma carpeta de este archivo. Están en archivo de excel o archivo de texto por si lo se les hace más fácil trabajar con ellos de una forma u otra. Cada archivo contiene datos de la calidad del aire en una ciudad por año. Cada fila está dividida por la primera columna que señala el día, la segunda la zona de la ciudad, la tercera el parámetro que se midió y las siguientes 24 son cada una la medición de dicho parámetro en cada hora del día.

Entregables:

* Gráficas

Es importante poder visualizar los datos con los que trabajaremos, como humanos nos es más sencillo entender figuras y colores que solo números. Utiliza la herramienta o lenguaje que gustes para ver el comportamiento de las variables y añádele de tu creatividad para sorprendernos con gráficas vistosas y coloridas. Grafica todo lo que puedas, antes y después de procesar los datos y el resultado de tu predicción, verás qué útil es.

* Limpieza de los datos

Los datos que te estamos pasando no son perfectos, pueden tener errores que al leerlos pueden perjudicar tu resultado final. Nunca podemos confiar en que nuestros datos son 100% certeros, habrá espacios vacíos, errores de medición o incluso alguna otra cosa que no tomamos en cuenta y tú descubrirás. Queremos que busques una forma de solucionarlo. Llena los espacios vacíos con el promedio del día anterior, ignora los valores nulos o utiliza cualquier forma que se te ocurra para limpiar los datos. ¡Solo recuerda no modificarlos demasiado porque puede afectar los resultados de tu predicción!

* Resultados de la predicción

Ahora sí, la parte buena. Cuentas con los datos de la calidad del aire de una ciudad desde 1996 hasta 2011, creemos que son suficientes datos para predecir los siguientes… ¿4 años? Desarrolla una regresión lineal, utiliza machine learning o ingéniatelas para crear tu propio modelo de predicción y demuéstrales a todos tus poderes de vidente. Al final, deberías tener archivos de texto con el mismo formato que los datos que te proporcionamos, pero de los años 2012, 2013, 2014 y 2015. No es necesario que lo hagas para todos los parámetros, tú decide cuales son los relevantes.

* Descripción de procesos y análisis [Formato libre] [Opcional] [BONUS]

Queremos que en un documento nos cuentes como le hiciste para resolver el reto, qué herramientas utilizaste para graficar, qué métodos de limpieza te ingeniaste y por qué, platícanos cómo le hiciste para predecir los cuatro años; pero, sobre todo, cuéntanos de tu análisis: ¿Qué patrones observaste?, ¿Notaste algo raro en los datos?, ¿Puedes decirnos cual es la zona más contaminada? ¿A qué crees que se deban los cambios de la calidad del aire? El documento es de formato libre, ponle tu estilo y creatividad.

Pitch: Una vez terminado el límite de entrega de proyectos los mentores visitarán tu mesa para que les expliques tu proyecto. Te pueden preguntar sobre cualquier entregable, proceso que utilizaste o que les muestres tu código así que prepárate para sorprenderlos. Después del pitch escogeremos a los 3 mejores equipos a que pasen al frente a presentar su proyecto. Cuéntales a todos lo que hiciste, muéstrales tus gráficas, fórmulas, resultados y descubrimientos, véndenos tu solución y demuestra por qué es la mejor. Te sugerimos crear una diapositiva, pero si lo prefieres o no te da tiempo pasa con tu laptop para que nos muestres a todos tu proyecto.

Si no sabes cómo proceder con alguno de los pasos señalados, ¡no te preocupes! sabemos que el reto puede ser difícil de completar, y para eso nos acompañan mentores de DataLab. Ellos estarán disponibles durante el evento para resolver cualquier duda. Además, tendremos talleres impartidos por ellos para ayudarte a completar cada una de las fases del reto. Mira la agenda del evento en codegdl.com, de igual forma daremos avisos para que no te los pierdas.

Para subir el proyecto es necesario que te registres en este link: [codegdl-18.devpost.com](http://codegdl-18.devpost.com)