

PROYECTO INTEGRADOS – GRUPO MMZMM4 (ALFA)

Miembros:

- Daniela Maidana
- Carolina Margheim
- Jorge Antonio Marchisone
- Mauricio Horacio Zenere

Proyecto “Pokemon Dataset”

Tomando como base para nuestro trabajo, la página web <https://pokemondb.net/pokedex/all>, nos enfocamos en extraer la información de cada personaje Pokemon, y sobre la dicha información, creas listas, informando cuales son los personajes mas potentes en ataque y defensa, mas rápidos y demás información que pudiéramos considerar relevante y que sea posible analizar.

Desarrollo de actividades y tareas. Desglose archivo ProyectoIntegrador.ipynb

Comenzamos importando las librerías necesarias sobre las que trabajamos.

En primer término, importamos la url desde la cual tomamos los datos necesarios para nuestra Base de Datos. Luego creamos listas y comenzamos a organizar los datos recolectados.

Con nuestras listas completas, creamos un DataFrame usando pandas, en donde definimos los nombres de cada columna, y su correspondiente lista.

Nuestro Data Set cuenta con un total de 1075 filas y 11 columnas, contamos la totalidad de los registros y chequeamos si tenemos datos nulos. Le pedimos a nuestro data frame que si encuentra un string vacío lo reemplace con un null ya que notamos que no todos los pokemones tienen la característica Tipo 2 (Ver archivo). De esta forma detectamos que para la variable 'Tipo 2', existen 495 registros vacíos. Sólo algunos Pokemones tienen una clasificación secundaria, por eso detectamos que existen 495 Pokemones sin Tipo 2 asignado. En total tenemos 1075 Pokemones.

Teniendo toda esta información procedimos a clasificar los personajes Pokemon y realizamos listas con los 5 personajes, con mayor puntaje para las Variables Defensa, Ataque, Velocidad de ataque y defensa. Velocidad y Vida.

También procedimos a clasificar los Pokemos de acuerdo a su Tipo 1 y Tipo 2.

Luego, a fin de realizar un análisis de correlación entre las variables, para ello, crearemos un nuevo data frame que solo cuenta con las variables numéricas para poder efectuar el análisis. Lo que nos permitió observar en el mapa de correlaciones las las características con la más alta correlación con respecto a Poder, diremos que Ataque, Ataque SP y Defensa SP tienen una muy alta correlación.

Algoritmos de Clasificación

KNN

Utilizamos scikit learn para trabajar con algoritmos de clasificación y agrupamiento. En primer lugar, extraemos sólo los datos que nos interesan y nos sirven a los fines de predecir la clasificación. Luego, entrenamos el algoritmo y obtenemos la predicción.

Este algoritmo de aprendizaje supervisado, utiliza la proximidad para hacer predicciones o clasificaciones de los nuevos datos partiendo del set de datos original.

Aplicando el modelo de KNN obtenemos una precisión del 17.34% en el set de test y de 45.35% en el set de entrenamiento.

REGRESSION

Continuamos utilizando la librería scikit learn. En primer lugar, obtenemos los valores únicos de la variable a predecir (Tipo 1) y a cada uno de ellos les asignamos un número de referencia ya que debemos trabajar con valores numéricos. Luego, creamos una instancia para la regresión (lr), entrenamos el algoritmo y obtenemos la predicción.

Este algoritmo de aprendizaje supervisado, utiliza la proximidad para hacer predicciones o clasificaciones de los nuevos datos partiendo del set de datos original.

El resultado de aplicar Regresión Lineal nos da un $R^2 = 0.04495$, lo que quiere decir que el modelo es poco fiable ya que se ajusta muy poco a la variable que estamos intentando explicar.