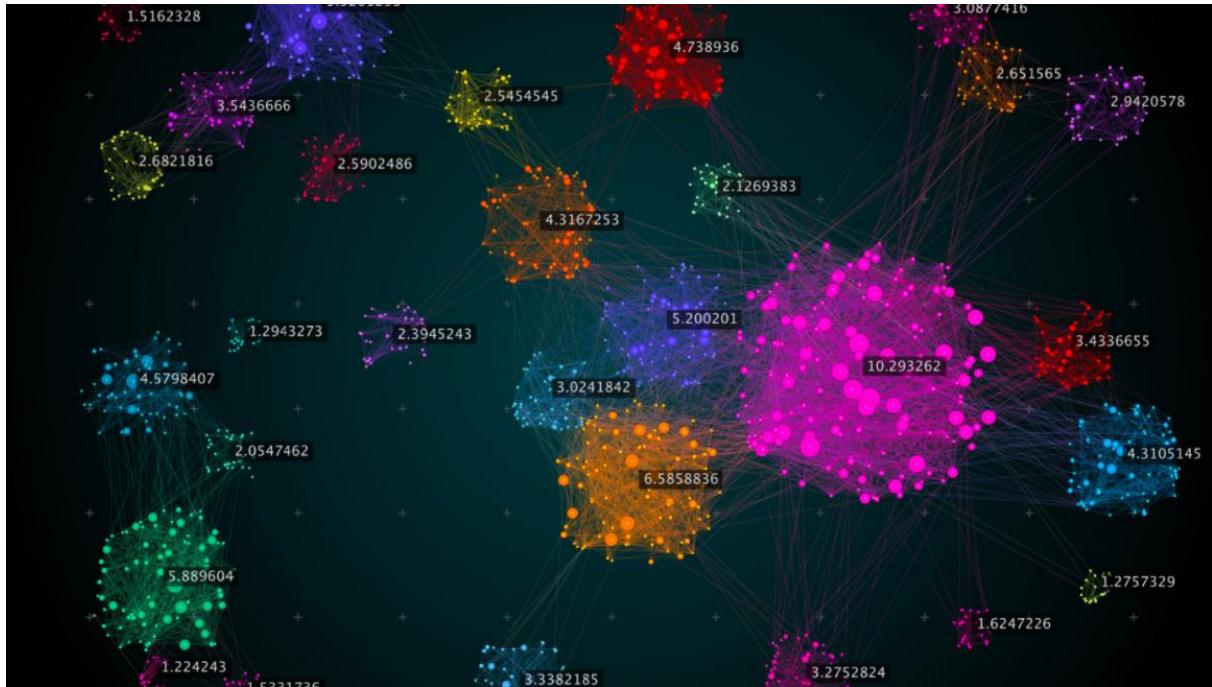


Trabajo práctico III – Agrupamiento

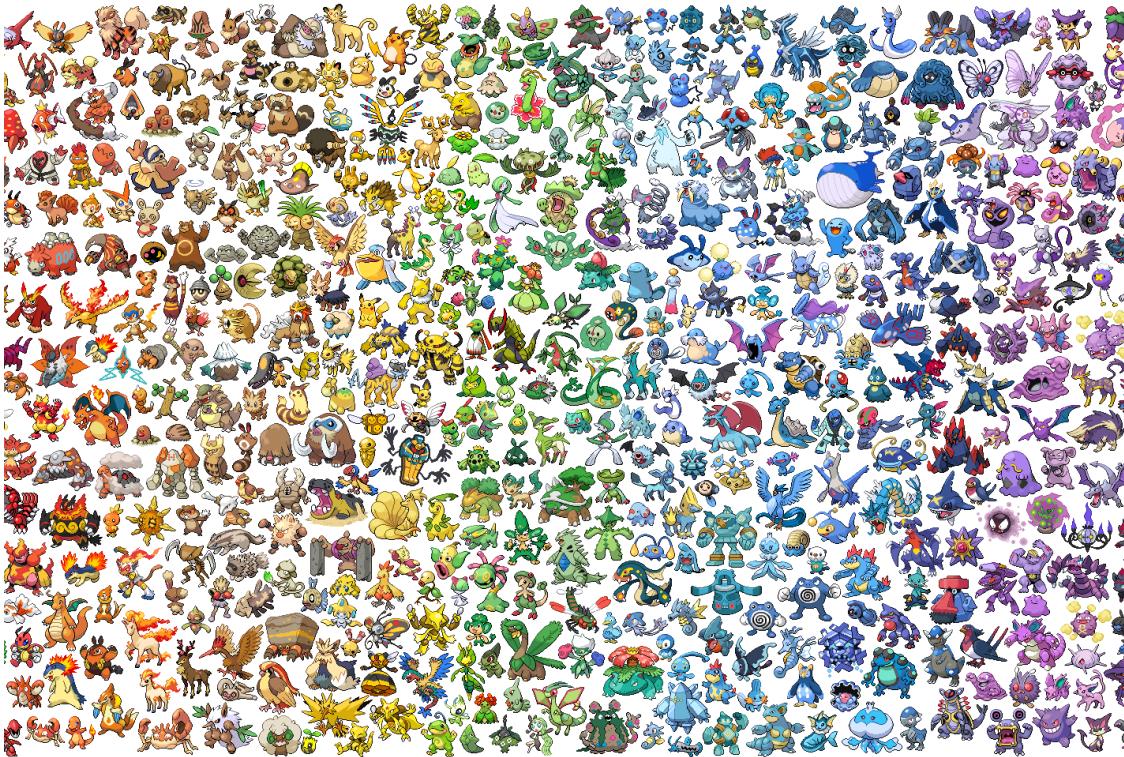


Requisitos básicos

- Jupyter para Python.
- Pandas.
- Numpy .
- Matplotlib y/ o nacido en el mar .

Práctico trabajar

El mundo Pokémon se ha convertido en un fenómeno masivo en todo el planeta, convirtiéndose en un campo de estudio que despierta un interés comparable al de la biología, la ecología e incluso la antropología moderna. Con cada nueva generación que introduce nuevas especies, hábitats y comportamientos, es cada vez más evidente que los jugadores se beneficiarían enormemente de una base de conocimientos más rigurosa.



Por ello, surge la idea de desarrollar un enfoque científico que permita comprender más profundamente los distintos grupos de Pokémon que existen, cómo se originan, cómo evolucionan, las relaciones que mantienen entre sí y las formas en las que interactúan con los ecosistemas que habitan.

Ejercicio 1 (Agrupamiento general)

El equipo de investigación aún no está seguro de si la tipificación Pokémon es la característica más estrechamente relacionada con rasgos ecológicos o evolutivos clave. Para investigar esto, buscan determinar si las características conductuales o fisiológicas observables se asocian significativamente con los diferentes tipos elementales.

Para explorar esta relación, se debe aplicar un algoritmo de agrupamiento para evaluar si los Pokémon se agrupan naturalmente de acuerdo con sus tipos oficialmente asignados. Como parte de este análisis, se deben realizar los siguientes pasos:

Paso 1 - Medio ambiente preparación

- Cree el notebook llamado **PW3_E1_GROUP[n].ipynb** (**n** es su número de grupo), en el servidor de notebook de su elección (Colab , Jupyter , etc .).
- Descargue el conjunto de datos de Pokémon del archivo [pokemons.json](#) [Descargar pokemons.json](#).

- Codifique nuestros datos para que cada Pokémon sea identificable de forma única y representado de una manera que permita mediciones de distancia significativas entre ellos.
- Explora el conjunto de datos:
 - Comprenda cada característica y su distribución.
 - Analizar las relaciones entre cada atributo y el objetivo.
 - Generar visualizaciones para respaldar la comprensión de estas.
- Preprocesar y transformar los datos para prepararlos para la regresión (si es necesario):
 - Característica selección o extracción.
 - Manejo desaparecido valores .
 - Codificar variables categóricas si es necesario.
 - Escalado o normalización si es necesario.

IMPORTANTE: Debes justificar las técnicas de limpieza y preprocesamiento que has aplicado.

Paso 2 - Implementación

- Aplique el algoritmo de agrupamiento K-Means utilizando diferentes valores de k para explorar cómo se agrupan naturalmente los datos.
- Evalúe qué valores de K producen los grupos más significativos y justifique la selección del k óptimo utilizando evidencia empírica.
- Aplique DBSCAN utilizando un rango de valores eps y min_samples para identificar regiones densas y posibles valores atípicos.
- Evalúe qué combinación de valores eps y min_samples produce los clústeres más significativos y justifique la selección de los valores óptimos (eps y min_samples) utilizando evidencia empírica.

No es necesario realizar búsquedas exhaustivas de hiperparámetros ni incluir todos los resultados de las pruebas en el informe. Simplemente incluya las mejores configuraciones de hiperparámetros que haya encontrado para cada técnica.

Paso 3 - Evaluación

- ¿Cómo se compara el modelo K-Means de mayor rendimiento con el modelo DBSCAN de mayor rendimiento en términos de estructura de

clústeres, interpretabilidad y capacidad para capturar patrones significativos en los datos de Pokémon? Analice las fortalezas y limitaciones de cada enfoque con base en sus resultados empíricos.

- Según su análisis, ¿qué técnica proporciona la agrupación más adecuada para este conjunto de datos? Justifique su elección con evidencia de ambos métodos .
- Responda las siguientes preguntas sobre el mejor modelo seleccionado:
 - ¿Los grupos generados por el algoritmo se alinean con los [tipos elementales oficiales de Pokémon](#)? , lo que sugiere que el tipo es de hecho un predictor significativo de patrones ecológicos o de comportamiento ?
 - ¿Qué características observables (por ejemplo, estadísticas, habilidades, variables de hábitat) parecen contribuir más fuertemente a la formación de grupos distintos, y cómo esto respalda o contradice las suposiciones existentes sobre la taxonomía de Pokémon?
 - Describe las características de tres grupos diferentes e intenta definir una etiqueta para ellos como en un problema de clasificación.

visualización en cada pregunta. Además, debe explicar por qué eligió esa visualización en particular.

Ejercicio 2 (Agrupamiento jerárquico)

En el estudio científico del mundo Pokémon, la clasificación tradicional basada en tipos elementales ofrece una visión general útil, pero no logra captar toda la complejidad biológica y funcional de estas criaturas. Aunque dos Pokémon pertenezcan a tipos diferentes, pueden desempeñar funciones similares, compartir estrategias de supervivencia o mostrar comportamientos similares en combate.

En este contexto, existe una necesidad de investigación que explore nuevas formas de clasificación basadas en características más profundas; específicamente, un estudio que emplee la agrupación jerárquica para analizar similitudes funcionales y construir una estructura relacional que refleje las verdaderas capacidades y patrones de comportamiento de los Pokémon.

Paso 1 - Medio ambiente preparación

- Reutilice el notebook anterior, si es posible, para crear uno nuevo llamado **PW3_E2_GROUP[n].ipynb** (**n** es su número de grupo), en el servidor de notebook de su elección (Colab , Jupyter , etc.).

Paso 2 - Implementación

- Aplicar agrupamiento jerárquico aglomerativo utilizando diferentes criterios de vinculación (único, completo, promedio, barrio).
- Definir un conjunto de métricas para elegir el mejor modelo.
- Genere un dendrograma para visualizar la estructura jerárquica para la mejor agrupación.

No es necesario realizar búsquedas exhaustivas de hiperparámetros ni incluir todos los resultados de las pruebas en el informe. Simplemente incluya las mejores configuraciones de hiperparámetros que haya encontrado para cada técnica.

Paso 3 - Evaluación

- Respuesta el próximo preguntas :
 - ¿Cómo se distribuyen los Pokémon (puntos de datos) en los diferentes grupos?
 - ¿Cuál considera que es el número óptimo de clústeres según la estructura del dendrograma? Justifique. su respuesta usando empírico evidencia .
 - ¿Los grupos resultantes revelan grupos funcionales que difieren de los tipos oficiales de Pokémon? cualquier inesperado relaciones o convergentes comportamientos.
 - ¿Zapdos , ZapdosGalar y Raichu se ubican en el mismo grupo o en grupos estrechamente relacionados? Interprete su ubicación en el dendrograma y explique qué indica esto sobre su similitud funcional según su conjunto de características.

visualización en cada pregunta. Además, debes explicar por qué la elegiste.

Documentación

Crea un documento de investigación, llamado **RESEARCH_PW3_GROUP[n].pdf** (**reemplaza [n] con tu número de grupo**) , resumiendo el trabajo práctico siguiendo la misma estructura:

- La portada incluye los nombres de todos los miembros del grupo junto con el número de grupo.

- Abstract o resumen.
- Tabla de Contenido .
- Introducción que describe un resumen conciso de los objetivos del PW, centrándose en su objetivo principal en lugar de copiar las instrucciones prácticas de trabajo.
- Información para cada ejercicio:
 - Descripción del conjunto de datos, incluida su estructura y los atributos clave utilizados para su solución.
 - Descripción de los métodos utilizados para la limpieza, preprocesamiento y visualización de datos, justificados por su propósito y relevancia.
 - Descripción de los experimentos realizados y sus respuestas a las diferentes preguntas.
- Conclusiones que resumen los principales resultados, aprendizajes e implicaciones de los experimentos.
- Referencias (si necesario).

Hay una plantilla de Word disponible en Canvas para [descargar](#), aunque su uso es opcional.

Normas

- Los trabajos prácticos deberán realizarse en grupos (3-4 personas).
- Los cuadernos se ejecutan correctamente y ejecutan todas las tareas requeridas en cada sección de la práctica. El profesor no evaluará el código en sí, sino que borrará todos los resultados generados en el cuaderno y ejecutará el programa desde cero para verificar su funcionalidad.
- Todos los resultados, tablas y figuras incluidos en el informe deben coincidir con los generados en el cuaderno.
- Todas las cifras incluidas en el informe deben permanecer claramente visibles en todos los niveles de detalle.
- **Código fuente Los sistemas de generación de contraseñas, como Github Copilot , GEMINI o ChatGPT , están estrictamente prohibidos.**
- Se deben entregar tres documentos separados: (1) un **documento de investigación** en formato PDF; y (2) dos cuadernos. Estos archivos deben

estar comprimidos en un archivo llamado PW3_GROUP[n].zip. (reemplace [n] con su número de grupo).

- El documento de investigación deberá tener un número máximo de **14 páginas. sin contar la portada, resumen, índice y bibliografía.**