**K-Nearest neighbors (kNN)**

**Ejemplo 1:** Hay los siguientes datos de flores del género Iris, donde cada flor está caracterizada por dos variables: ***Sepal Length*** y ***Sepal Width***. Además, se conoce la especie de cada flor: **Setosa**, **Versicicolor** y **Virginica**.

El objetivo es predecir la especie de una nueva flor desconocida basándose en la similitud con las flores existentes en el conjunto de datos usando el algoritmo de *K-Nearest Neighbors* (*kNN*) para predecir la especie de la nueva flor.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sepal Length** | **Sepal Width** | **Specie** | **k=1** | **k=3** |
| *Nueva flor* | 5.9 | 2.8 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sepal Length** | **Sepal Width** | **Species** | **Distancia euclidiana** | **Ranking** |
| 5.3 | 3.7 | Setosa |  |  |
| 5.1 | 3.8 | Setosa |  |  |
| 7.2 | 3.0 | Virginica |  |  |
| 5.4 | 3.4 | Setosa |  |  |
| 5.1 | 3.3 | Setosa |  |  |
| 5.4 | 3.9 | Setosa |  |  |
| 7.4 | 2.8 | Virginica |  |  |
| 6.1 | 2.8 | Versicicolor |  |  |
| 7.3 | 2.9 | Virginica |  |  |
| 6.0 | 2.7 | Versicicolor |  |  |
| 5.8 | 2.8 | Virginica |  |  |
| 6.3 | 2.3 | Versicicolor |  |  |
| 5.1 | 2.5 | Versicicolor |  |  |
| 6.3 | 2.5 | Versicicolor |  |  |
| 5.5 | 2.4 | Versicicolor |  |  |

1. Realiza el gráfico de dispersión de los datos obtenidos.
2. Calcula la distancia euclidiana entre la nueva flor y cada flor del conjunto de datos y clasifícalas de menor a mayor distancia.
3. ¿Cuál sería la especie de la nueva flor si o ? Razona la respuesta.

