Tugas : Membuat Rule Deskriptif Menggunakan Bantuan Tools Excel Untuk Mengelompokkan Jenis Bunga Iris

Nama : Ayuningtyas Hari Fristiana

Kelas : Big Data – B

**Rule untuk mengelompokan jenis iris adalah sebagai berikut:**

1. **Jika 0.1<= petal\_width<=0.6 dan 1<=petal\_length<=1.9 maka termasuk Iris-setosa**
2. **Jika 1<= petal\_width<=1.8 dan 3<=petal\_length<=5.1maka termasuk Iris-versicolor**
3. **Jika 1.9<= petal\_width<=2.5 dan 4.5<=petal\_length<=6.9 maka termasuk Iris-virginica**

Penjelasannya adalah sebagai berikut:

Pivot Table Range Untuk Petal Length dan Width

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Min of petal\_length** | **Max of petal\_length** | **Min of petal\_width** | **Max of petal\_width** |
| Iris-setosa | 1 | 1.9 | 0.1 | 0.6 |
| Iris-versicolor | 3 | 5.1 | 1 | 1.8 |
| Iris-virginica | 4.5 | 6.9 | 1.4 | 2.5 |
| **Grand Total** | **1** | **6.9** | **0.1** | **2.5** |

Pivot Table Range Untuk Sepal Length dan Width

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Min of sepal\_length** | **Max of sepal\_length** | **Min of sepal\_width** | **Max of sepal\_width** |
| Iris-setosa | 4.3 | 5.8 | 2.3 | 4.4 |
| Iris-versicolor | 4.9 | 7 | 2 | 3.4 |
| Iris-virginica | 4.9 | 7.9 | 2.2 | 3.8 |
| **Grand Total** | **4.3** | **7.9** | **2** | **4.4** |

Berdasarkan pegamatan dari pivot tabel tersebut terlihat bahwa yang mempengaruhi pengelompokkan jenis bunga iris adalah ada pada Petal length atau width. Untuk sepal tidak signifikan berpengaruh. Untuk lebih jelas mari kita lihat pada scatter plot berikut ini:

Iris-setosa dilambangkan bentuk kotak warna merah

Iris-versicolor dilambangkan bentuk diamond warna hijau

Iris-virginica dilambangkan bentuk segitiga warna biru

Dari scatter plot di bawah dapat dilihat pula bahwa variable petal signifikan pengaruhnya terhadap pengelompokkan jenis Iris. Selanjutnya mari kita amati lebih lanjut apakah length atau width yang lebih significant.

Kalau kita amatai terlihat masih ada area yang berbaur antara versicolor dengan virginica. Oleh karena itu kita lakukan pengamatan outlier untuk mengetahui apakah yang berbaur tersebut adalah outlier dari kelompok mana.

Dari berbagai scenario scatter plot terlihat bahwa yang lebih signifikan adalah variable Petal Width di mana ada 5 varian virginica yang merupakan outlier bercampur dengan varian versicolor.

Sehingga untuk mendapatkan pengelompokkan yang lebih baik maka 5 outlier tersebut akan dikeluarkan dari dataset yaitu varian iris-virginica yang petal width yang kurang dari 1.8. Dan berikut hasil scatter plot setelah dilakukan cleansing data.



Sehingga Rule untuk mengelompokan jenis iris adalah sebagai berikut:

1. Jika 0.1<= petal\_width<=0.6 dan 1<=petal\_length<=1.9 maka termasuk Iris-setosa
2. Jika 1<= petal\_width<=1.8 dan 3<=petal\_length<=5.1maka termasuk Iris-versicolor
3. Jika 1.9<= petal\_width<=2.5 dan 4.5<=petal\_length<=6.9 maka termasuk Iris-virginica