18-12-2024

Miguel Ángel Lozoya Blázquez

Jorge Galán Rodríguez

Feedback1-parte1

Análisis Vinos

(aprendizaje no supervisado)

**Botella y copa de vino

Descripción generada automáticamente**

1. **MODELO** 
   1. **Valores estadísticos:**

**Vinos blancos**

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

**Vinos tintos**

**Pantalla de computadora

Descripción generada automáticamente con confianza media**

Primero, se ha hecho un describe() para ver los valores estadísticos como media, la desviación estándar, valores mínimos y máximos para los vinos blancos y tintos.

**VALORES IDEALES**

* **Fixed Acidity :**
  + Vinos Blancos entre 6.0 –8.5 g/L
  + Vinos Tintos entre 6.5 – 10.0 g/L
* **Volatile Acidity:**
  + Vinos Blancos ≤ 0.35 g/L
  + Vinos Tintos ≤ 0.60 g/L
* **Citric Acid:**
  + Vinos Blancos entre 0.2 – 0.5 g/L
  + Vinos Tintos entre 0.0 – 0.3 g/L
* **Residual Sugar;**
  + Vinos Blancos:
    - Seco ≤ 4g/L
    - Semiseco entre 4 – 12 g/L
    - Dulce > 45g/L
  + Vinos Tintos:
    - Seco ≤ 4g/L
    - Es extraño verlos con niveles más altos a 4g/L
* **Chlorides:** 
  + Vinos Blancos entre 0.02 – 0.06 g/L
  + Vinos Tintos entre 0.04 – 0.09 g/L
* **Free Sulfur Dioxide:**
  + Vinos Blancos entre 20 – 50 mg/L
  + Vinos Tintos entre 10 – 35 mg/L
* **Total Sulfur Dioxide:** 
  + Vinos Blancos ≤ 200 mg/L
  + Vinos Tintos ≤ 150 mg/L
* **Density:** 
  + Vinos Blancos entre 0.990 – 0.996 g/cm3
  + Vinos Tintos entre 0.995 – 1.000 g/cm3
* **pH:** 
  + Vinos Blancos entre 3.0 – 3.4
  + Vinos Tintos entre 3.3 – 3.6
* **Sulphates:**
  + Vinos Blancos entre 0.4 – 0.6 g/L
  + Vinos Tintos entre 0.5 – 0.8 g/L
* **Alcohol:**
  + Vinos Blancos entre 9 – 13% vol
  + Vinos Tintos entre 11-15% vol
* Si nos fijamos en los valores count, vemos que no hay ningún valor faltante.
* En el promedio, vemos lo siguiente:

**Vinos blancos**

* + **Fixed acidity:**
    - **Promedio** es de 6.85. Esto nos dice que una acidez más baja que afecta que sean más frescos y ligeros.
    - **Desviación estándar** es de 0.83 que nos indica que hay baja dispersión en los datos.
    - **Percentiles** 
      * **Mínimo:** 3.8 puede ser outlier
      * **25% (Q1):** 6.3 es un valor típico.
      * **Mediana (50%):** 6.8 (nivel del blanco seco)
      * **75% (Q3):** 7.3 (nivel alto, mayor estructura)
      * **Máximo:** 14.2 (por encima del valor lógico, posible outlier, valor erróneo o atípico)
  + **Volatile acidity:**
    - **Promedio** es de 0.28. Valores altos puede afectar en el sabor.
    - **Desviación estándar** es de 0.10. Al estar cerca del promedio, nos indica que hay poca dispersión.
    - **Percentiles** 
      * **Mínimo:** 0.08 (muy bajo)
      * **25% (Q1):**.0.21 (bajo)
      * **Mediana (50%):** 0.26
      * **75% (Q3):** 0.32
      * **Máximo:** 1.10 (es un outlier)
  + **Citric Acid:** 
    - el promedio es de 0.33. Esto afecta a la frescura del vino.
    - **Desviación estándar** 0.12, indica baja dispersión.
    - **Percentiles** 
      * **Mínimo:** 0.0 (es un valor más bajo de lo normal, ¿outlier?)
      * **25% (Q1):**. 0.27
      * **Mediana (50%):** 0.32
      * **75% (Q3):** 0.39
      * **Máximo** : 1.66 (es un outlier)
  + **Residual Sugar:** 
    - el promedio es de 6.40. Es normal que tengan un valor más alto, porque los blancos son más dulces.
    - **Desviación estándar** es de 5.11, indica una alta dispersión de los datos.
    - **Percentiles** 
      * **Mínimo:** 0.6 (sería un posible valor para vino semi seco)
      * **25% (Q1):** 1.7 (vino seco)
      * **Mediana (50%):** 5.2blancos semisecos
      * **75% (Q3):** 9.9 indica semi seco
      * **Máximo** : 65.8 vinos muy dulces (revisar que no sea outlier)
  + **Chlorides:** 
    - el promedio es de 0.045. Es un nivel bajito que afecta al sabor.
    - **Desviación estándar** es de 0.02, indica baja dispersión
    - **Percentiles** 
      * **Mínimo:** 0.0009 (muy bajo, revisar que no sea un outlier)
      * **25% (Q1):** 0.035
      * **Mediana (50%):** 0.043
      * **75% (Q3):** 0.050
      * **Máximo** : 0.29
  + **Free Sulfur Dioxide:**
    - el promedio es de 35.37. Esto sirve para proteger el vino de la oxidación.
    - **Desviación estándar** es de 17.12, valor muy variados.
    - **Percentiles** 
      * **Mínimo:** 2 (es un outlier)
      * **25% (Q1):**. 23
      * **Mediana (50%):** 34
      * **75% (Q3):** 46
      * **Máximo** : 289 (puede ser un error en los datos)
  + **Total Sulfur Dioxide:** 
    - el promedio es de 138.33. Es alto, pero optimo para la preservación.
    - **Desviación estándar** es de 42.52, indica alta dispersión.
    - **Percentiles** 
      * **Mínimo:** 10
      * **25% (Q1):**. 108
      * **Mediana (50%):** 134
      * **75% (Q3):** 167
      * **Máximo** : 440 (es un outlier)
  + **Density:** 
    - el promedio es de 0.994. Es bajo, pero es un valor normal en los vinos blancos.
    - **Desviación estándar** es de 0.003, indica baja dispersión de los datos.
    - **Percentiles** 
      * **Mínimo:** 0.987
      * **25% (Q1):**. 0.991
      * **Mediana (50%):** 0.994
      * **75% (Q3):** 0.996
      * **Máximo** : 1.038 (este valor puede ser un vino dulce)
  + **pH:** 
    - el promedio es de 3.19. Esto significa que el pH es ácido, pero es un valor normal en los blancos.
    - **Desviación estándar** es de 0.15, indica una baja dispersión.
    - **Percentiles** 
      * **Mínimo:** 2.72 (posible outlier, dato erroneo, valor atípico)
      * **25% (Q1):**. 3.08
      * **Mediana (50%):** 3.17
      * **75% (Q3):** 3.28
      * **Máximo** : 3.81 (posible outlier, dato erroneo, valor atípico)
  + **Sulphates:** 
    - el promedio es de 0.49. Es un valor correcto, ayuda a estabilizar el vino.
    - **Desviación estándar** es de 0.11, indica baja dispersión
    - **Percentiles** 
      * **Mínimo:** 0.22 (posible outlier)
      * **25% (Q1):**.0.41
      * **Mediana (50%):** 0.48
      * **75% (Q3):** 0.55
      * **Máximo** : 1.08 (posible outlier)
  + **Alcohol:** 
    - el promedio es de 10,52%. Es un valor dentro del rango correcto.
    - **Desviación estándar** es de 1.23 bastante variabilidad
    - **Percentiles** 
      * **Mínimo:** 8.0% (posible outlier)
      * **25% (Q1):** 9,5%
      * **Mediana (50%):** 10.4%
      * **75% (Q3):** 11.4%
      * **Máximo** : 14.2% (posible outlier o dato erroneo)

**Vinos tintos**

* + **Fixed acidity:**
    - el promedio es de 6.85. Esto nos dice que una acidez más baja que afecta que sean más frescos y ligeros.
    - **Desviación estándar** es de 1.73, mayor dispersión.
    - **Percentiles** 
      * **Mínimo:** 4.6 (posible outlier)
      * **25% (Q1):**.7.1
      * **Mediana (50%):** 7.9
      * **75% (Q3):** 9.2
      * **Máximo** : 15.9 (posible outlier)
  + **Volatile acidity:**
    - el promedio es de 0.28. Valores altos puede afectar en el sabor.
    - **Desviación estándar** es de 0.18**,** indica baja dispersión de los datos.
    - **Percentiles** 
      * **Mínimo:** 0.12
      * **25% (Q1):**.0.40
      * **Mediana (50%):** 0.52
      * **75% (Q3):** 0.64
      * **Máximo** : 1.58 (posible outlier)
  + **Citric Acid:** 
    - el promedio es de 0.33. Esto afecta a la frescura del vino.
    - **Desviación estándar** es de 0.19, dispersión de los datos.
    - **Percentiles** 
      * **Mínimo:** 0.0
      * **25% (Q1):**.0.09
      * **Mediana (50%):** 0.25
      * **75% (Q3):** 0.42 (posible outlier)
      * **Máximo** : 1.0 (posible outlier)
  + **Residual Sugar:** 
    - el promedio es de 6.40. Es normal que tengan un valor más alto, porque los blancos son más dulces.
    - **Desviación estándar** es de 1.40, indica dispersión de datos.
    - **Percentiles** 
      * **Mínimo:** 0.09
      * **25% (Q1):**. 1.9
      * **Mediana (50%):** 2.2
      * **75% (Q3):** 2.6
      * **Máximo** : 15.5 (posible error o outlier)
  + **Chlorides:** 
    - el promedio es de 0.045. Es un nivel bajito que afecta al sabor.
    - **Desviación estándar** es de 0.047, indica baja dispersión.
    - **Percentiles** 
      * **Mínimo:** 0.012 (outlier, valor atípico)
      * **25% (Q1):**.0.070
      * **Mediana (50%):** 0.079
      * **75% (Q3):** 0.090
      * **Máximo** : 0.611 (es un outlier)
  + **Free Sulfur Dioxide:**
    - el promedio es de 35.37. Esto sirve para proteger el vino de la oxidación.
    - **Desviación estándar** es de 10.42, mucha variedad en los datos.
    - **Percentiles** 
      * **Mínimo:** 1.0 (posible outlier, valor atípico)
      * **25% (Q1):**.7.0
      * **Mediana (50%):** 13.0
      * **75% (Q3):** 21.0
      * **Máximo** : 72.0 (posible outlier, valor atípico)
  + **Total Sulfur Dioxide:** 
    - el promedio es de 138.33. Es alto, pero optimo para la preservación.
    - **Desviación estándar** es de 33.22, indica alta dispersión
    - **Percentiles** 
      * **Mínimo:** 6.0
      * **25% (Q1):**.22.0
      * **Mediana (50%):**38.0
      * **75% (Q3):** 62.0
      * **Máximo** : 289 (outlier, valor atípico)
  + **Density:** 
    - el promedio es de 0.994. Es bajo, pero es un valor normal en los vinos blancos.
    - **Desviación estándar** es de 0.0018, indica estabilidad en los datos.
    - **Percentiles** 
      * **Mínimo:** 0.990
      * **25% (Q1):**. 0.995
      * **Mediana (50%):** 0.996
      * **75% (Q3):** 0.997
      * **Máximo** : 1.036 (posible outlier, valor atípico)
  + **pH:** 
    - el promedio es de 3.19. Esto significa que el pH es ácido, pero es un valor normal en los blancos.
    - **Desviación estándar** es de 0.15, indica baja dispersión
    - **Percentiles** 
      * **Mínimo:** 2.74 (posible outlier, valor atípico)
      * **25% (Q1):**. 3.21
      * **Mediana (50%):** 3.31
      * **75% (Q3):** 3.40
      * **Máximo** : 4.01(posible outlier, valor atípico)
  + **Sulphates:** 
    - el promedio es de 0.49. Es un valor correcto, ayuda a estabilizar el vino.
    - **Desviación estándar** es de 0.17, indica una dispersión moderada.
    - **Percentiles** 
      * **Mínimo:** 0.37 (posible outlier, valor atípico)
      * **25% (Q1):**.0.55
      * **Mediana (50%):** 0.62
      * **75% (Q3):** 0.73
      * **Máximo** : 2.0 (posible outlier)
  + **Alcohol:**
    - el promedio es de 10,52%. Es un valor dentro del rango correcto.
    - **Desviación estándar** es de 1.07, indica una baja dispersión.
    - **Percentiles** 
      * **Mínimo:** 8.4% (posible outlier, valor atípico)
      * **25% (Q1):**. 9.5%
      * **Mediana (50%):** 10.1%
      * **75% (Q3):** 11.1%
      * **Máximo** : 14.9%

**Conclusiones:**

Con esto concluimos que debemos fijarnos en las columnas siguientes por posibilidad de outliers:

* Blancos : Residual Sugar, Total Sulfur Dioxide, Free Sulfur Dioxide.
* Tintos: Residual Sugar, Free Sulfur Dioxide, Volatile Acidity.

**1.2 Boxplot para identificar outliers o anomalías:**

**Vinos blancos**

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamenteGráfico, Gráfico de cajas y bigotes

Descripción generada automáticamenteGráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

**Conclusión:** estos boxplots nos ha servido para hacernos una idea de como estaban distribuidas las características.

Podemos decir que, las características como citrid acid, volatile acidity y fixed acidity, tienes los valores bastante equilibradas, aunque presentan algún valor atípico, puede ser totalmente lógico por alguna tipología de vino distinto.

Respecto a valores atípicos, podemos destacacar el residual sugar. Total sulfur dioxide y free sulfur dioxide y sulphates tiene variaciones más significativas, eso no indica que debemos prestar atención.

**Vinos tintos**

**Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente** **Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente** Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

**Conclusiones:** En el caso de los vinos tintos hemos visto que cumples las mismas características equilibradas (como citrid acid, volatile acidity y fixed acidity).

En el lado de las anomalías son bastante similares (residual sugar. Total sulfur dioxide y free sulfur dioxide y sulphates), salvo que en el caso del vino tinto hay que añadirle el alcohol que hay casos con un volumen de alcohol elevado para ese tipo de vino.

* 1. **Mapa de calor de correlación:**

Hemos utilizado está técnica para detectar las relación o dependencia entre variables.

**Vinos blancos**

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Nos resalta la relación entre:

Relación positiva:

* la density y el Residual Sugar con un 0.84. A medida que incrementa el Residual Sugar, también lo hace la Density.
* El Total Sulfur Dioxide y el Free Sulfur Dioxide con un 0.62 y tiene lógica por ser conservantes.

Relación negativa:

* El pH y Fixed Acidity con un -0.42. Esto nos indica que el pH es un indicador de la acidez del vino indirectamente.
* Alcohol y la Density con un -0.78. Esto nos indica que cuanto mayor es el volumen en alcohol del vino, menor es la densidad de este.
* Alcohol y Residual con un -0.45. Esto nos indica que cuanto mayor volumen en alcohol tiene el vino, menor es el azúcar residual de este, porque se ha consumido en el proceso de fermentación.

**Vino tinto**

Gráfico, Gráfico de barras, Gráfico en cascada

Descripción generada automáticamente

Nos resalta la relación entre:

Relación positiva:

* El Fixed Acidity y la Density con un 0.67. Esto nos indica que a medida que incrementa la acidez fija del vino, incrementa la densidad de éste.
* El Total Sulfur Dioxide y el Free Sulfur Dioxide con un 0.66. Nos indica una relación entre estas variables, porque el Free Sulfur Dioxide es una parte del Total Sulfur Dioxide.
* El Citric Acid y el Fixed Acidity con un 0.68. Esto tiene sentido porque el ácido cítrico es un componente significativo en la parte de acidez fija en los vinos tintos.

Relación negativa:

* El pH y Fixed Acidity con un -0.68. Al igual que en el vino blanco el pH es un indicador indirecto de la acidez fija del vino.
* Alcohol y la Density con un -0.49. Esto nos indica que cuanto mayor es el volumen en alcohol del vino, menor es la densidad de este.

**1.4 Reducción de dimensionalidad con PCA (Análisis de Componentes Principales):**

Aplicamos primero el Standard Scaler para que todas las variables tengan el mismo peso inicial antes de aplicar el PCA.

**Vinos blancos**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Esta técnica la hemos utilizado para saber el nº de componentes de que podemos utilizar que mejoraría el rendimiento en el modelo sin perder demasiada información. En este caso usaremos 7 componentes.

Posteriormente calculamos la distancia entre los k-vecinos y lo aplicamos:

Gráfico

Descripción generada automáticamente

A simple vista vemos que la distancia k-vecinos ronda el 2.

**Vinos tintos**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

En el caso de los vinos tintos hemos escogido 7 componentes.

Posteriormente calculamos la distancia entre los k-vecinos y lo aplicamos:

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

A simple vista vemos que en el caso de los vinos tintos rondaría en una distancia k-vecinos debería ser inferior a 2.

**1.5 Optimización de parámetro de DBSCAN:**

**Vinos blancos**

Tras los análisis del punto anterior hemos configurado las métricas y aplicamos las métricas de Silhouette y Calinski-Harabasz para ver la mejor combinación.

Pantalla de computadora

Descripción generada automáticamente con confianza media

En este caso hemos obtenido que los parámetros que deberemos utilizar para el DBSCAN son de eps=2.0 y min\_samples=13.

El coeficiente de Silueta es del 0.48 que nos sirve para saber que tal están las agrupaciones de los clústers que puede hacer el DBSCAN. Lo que indica que las agrupaciones están bastante bien y se pueden separar de otros clústers, pero aun faltaría mejorar la calidad.

El índice de Calinski-Harabasz nos ayuda a ver la dispersión de los clústers y la distancia entre ellos. El 107.01 nos dice que están bastante bien agrupados, pero es mejorable.

Ambos pueden estar afectados por el ruido.

**Vinos tintos**

Pantalla negra con letras blancas

Descripción generada automáticamente

En este caso hemos obtenido que los parámetros que deberemos utilizar para el DBSCAN son de eps=1.9 y min\_samples=9.

El coeficiente de silueta es de 0.59 que es algo mayor al de los vinos blancos por lo que podría agrupar un poco mejor esta tipología de vinos, pero es mejorable.

El índice de Calisnki-Harabasz nos ha dado 82.97 que eso es indicativo de que los clústers son más dispersos.

**1.6 Visualización de clusters y outliers**

**Vinos blancos**

Pasamos el DBSCAN para encontrar el número de outliers y sus valores.

****

Visualizamos ese clúster conseguido y los marcamos en rojo los valores outliers en esta representación en 2D.

**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

****

Se ha conseguido tener un resultado de valores dentro del clúster de 3757.

**Vinos tintos**

Con los vinos tintos hemos aplicado lo mismo que los vinos blancos, pero esta tipología de vinos hay más outliers.



Gráfico, Gráfico de dispersión

Descripción generada automáticamente



Por consiguiente nos quedamos con unos valores agrupados de 1181.

**1.7 Método del Codo sin Outliers**

Volvemos a hacer el método del codo, pero ahora le hemos quitado esos valores outliers.

**Vinos blancos**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Lo vemos representado es que a partir 4/5 clústers se empieza a aplanar la gráfica. Podemos concluir a primera vista que al eliminar los valores atípicos, reduciremos la dispersión y mejoraremos la agrupación de nuestros datos.

**Vinos tintos**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

En el caso de los vinos tintos, la gráfica se empieza a aplanar en 5/6 clústers.

**1.8 Silhouette utilizando K-means**

Utilizamos k-means para hacer la agrupación de los clústeres con la representación gráfica de Silhouette para ver que tan bien se ha hecho la asignación de los valores para llevarlos a un clúster u otro:

**Vinos blancos**

Gráfico

Descripción generada automáticamente

Como nos ha dado un índice Silhouette promedio no ha dado inferior a 0.2, entonces, hemos probado con 6 clusters o con 4 clústers para ver si mejoran nuestros resultados.

Gráfico

Descripción generada automáticamente

Con 4 clústers sigue teniendo el promedio bastante bajo, pero reducimos los puntos negativos por lo que hemos conseguido mejorar la calidad de los agrupamientos.

Gráfico

Descripción generada automáticamente

Escoger 6 clústers empeoraría mucho los resultados de nuestro agrupamiento.

**Vinos tintos**

En el caso de los vinos tintos hemos hecho el análisis con 5, 6 o 7 clústers.

Gráfico

Descripción generada automáticamente

Gráfico, Gráfico de embudo

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

El caso de 7 clústeres la hemos descartado por que baja significativamente el promedio del índice Silueta y eso implica que la calidad de la agrupamiento está empeorando.

Posteriormente, hemos comparado el de 5 y 6 clústers. Ambos tiene un promedio de índice muy similar, pero si nos fijamos en si escogemos la opción de 5 clústers la agrupación de los valores tiene una mejora.

**1.9 PCA utilizando K-means**

Por último hemos utilizado el PCA con el k-means para ver en 2D las agrupaciones con la selección que hemos visto en la gráfica Silhouette:

**Vinos blancos**

**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Finalmente nos hemos decantado por el uso de 4 clústers.

Interfaz de usuario gráfica

Descripción generada automáticamente

**Conclusiones de los resultados obtenidos:**

Clúster 0 (vinos suaves y equilibrados):

* Fixed acidity (6.41) y Acid Citric (0.30) tiene un nivel bajo.
* Total Sulfur Dioxide (136) es moderado.
* Alcohol (10.53) ligeramente bajo.
* pH (3.31) es alto.

Clúster 1 (vinos más ácidos):

* Fixed acidity (7.48) y Acid Citric (0.36) un poco más elevado.
* Total Sulfur Dioxide (115) es bajo.
* Alcohol (10.78) es moderado.
* Sulphates (0.45) y pH (3.09) es más bajos.

Clúster 2 (vino dulce, denso, pero con menor alcohol):

* Residual sugar (11.71) es significativamente bajo.
* Total Sulfur Dioxide (174) es mas elevado.
* Density (0.997) es alto.
* Alcohol (9.45) es más bajo.

Clúster 3 (vinos con alto volumen de alcohol y con buen aroma):

* Volatile Acid (0.34) es alto.
* Alcohol (11.95) es alto
* Total Sulfur Dioxide (108) es bajo.
* Sulfur (0.44) y pH (3.23) moderado.

Gráfico

Descripción generada automáticamente

Ejemplo de clasificación:

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

**Vinos tintos**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Finalmente nos hemos decantado por 5 clústers:

Imagen que contiene tabla, grande, estacionado, monitor

Descripción generada automáticamente

Clúster 0 (vino equilibrado):

* Fixed Acidity (8.23) y Citric Acid (0.26) son moderados.
* Sulphates (0.67) son los más altos.
* Alcohol (10.42) es moderado.
* Total Sulfur Dioxide (47.15) es muy bajo.

Clúster 1 (vino con alto volumen alcohol y textura robusta):

* Fixed Acidity (8.15) alta.
* Alcohol (10.95) es el más alto entre los grupos.
* Sulphates (0.64) elevados.
* Density (0.9975) también alta.

Clúster 2 (vinos entre dulces y semidulces):

* Residual Sugar (10.38) es muy alto, lo que indica vinos dulces.
* Total Sulfur Dioxide (283.56) es el más elevado.
* Alcohol (9.68) es bajo.
* Density (1.000) muy alta.

Clúster 3 (vino ácido con notas aromáticas):

* Volatile Acid (0.56) es la más alta, indicando posible carácter aromático intenso.
* pH (3.23) es el más bajo.
* Alcohol (9.81) también bajo.
* Total Sulfur Dioxide (126.46) moderado.

Clúster 4(vino con alto volumen alcohol y textura robusta):

* Fixed Acidity (8.47) alta con niveles equilibrados de ácido cítrico (0.26).
* Alcohol (10.62) alto.
* Sulphates (0.64) también elevados.
* Total Sulfur Dioxide (20.49) son bajos.

Gráfico

Descripción generada automáticamente

Ejemplo de clasificación del modelo:

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media