

Módulo: ADM1929 - Business understanding: Pensamiento analítico basado en datos. - (A51)

**Actividad: Reto de aprendizaje 12.
Ejercicio de Auto Model**

Nombre: Roberto Mora Balderas

Asesor: José Carlos Soto Monterrubio

Fecha: 21 de julio de 2023

Objetivo:

Aplicar los conocimientos obtenidos para generar el primer algoritmo predictivo con Auto Model.

Instrucciones:

Si bien este no es el módulo en el que nos concentraremos en realizar modelos predictivos a detalle, es importante que vayas viendo lo que puede hacer Rapid Miner.

1. Con el [data set](#) de la actividad anterior, repite el modelo de regresión lineal que elaboré en pantalla, intenta interpretar los resultados a partir del video en el que te explico en qué consiste. No te pido que busques mejorar el modelo de regresión, solo génalo y observa los valores resultantes. Pon más atención en el simulador generado, analiza el simulador y desarrolla un análisis en un documento de Word.

Desarrollo:

Se replicó el ejercicio y se optimizó de la misma manera obteniendo los siguientes resultados.

Generalized Linear Model - Model

| Attribute | Coefficient | Std. Coefficient | Std. Error | z-Value | p-Value |
|----------------------|-------------|------------------|------------|---------|---------|
| fixed.acidity | 0.005 | 0.008 | 0.034 | 0.142 | 0.887 |
| volatile.acidity | -0.910 | -0.159 | 0.162 | -5.603 | 0.000 |
| citric.acid | -0.022 | -0.004 | 0.191 | -0.116 | 0.907 |
| residual.sugar | 0.006 | 0.008 | 0.018 | 0.306 | 0.760 |
| chlorides | -1.874 | -0.078 | 0.606 | -3.092 | 0.002 |
| free.sulfur.dioxide | 0.009 | 0.099 | 0.003 | 3.351 | 0.001 |
| total.sulfur.dioxide | -0.005 | -0.153 | 0.001 | -4.917 | 0.000 |
| density | -6.561 | -0.012 | 28.288 | -0.232 | 0.817 |
| pH | -0.530 | -0.080 | 0.252 | -2.099 | 0.036 |
| sulphates | 0.868 | 0.148 | 0.146 | 5.957 | 0.000 |
| alcohol | 0.272 | 0.284 | 0.034 | 8.097 | 0.000 |
| Intercept | 11.164 | 5.611 | 27.677 | 0.403 | 0.687 |

Podemos observar aquí los siguientes puntos:

1. Signo e Impacto de los Coeficientes:

- Coeficientes positivos (por ejemplo, fixed.acidity, residual.sugar, free.sulfur.dioxide, sulphates, alcohol) indican que un aumento en el atributo se asocia con un aumento en la variable objetivo.
- Coeficientes negativos (por ejemplo, volatile.acidity, chlorides, total.sulfur.dioxide, pH) indican que un aumento en el atributo se asocia con una disminución en la variable objetivo.

2. Significancia de los Coeficientes:

- Los valores p proporcionan una indicación de la significancia de cada coeficiente. Un valor p pequeño (típicamente menor a 0.05) sugiere que el coeficiente es estadísticamente significativo en la predicción de la variable objetivo.
- Atributos con valores p muy pequeños (por ejemplo, volatile.acidity, chlorides, free.sulfur.dioxide, sulphates, alcohol) son probablemente importantes para predecir la variable objetivo.
- Atributos con valores p más grandes (por ejemplo, fixed.acidity, citric.acid, residual.sugar, density, pH) pueden no tener un impacto significativo en la variable objetivo en este modelo.

3. Coeficientes Estandarizados:

- Los coeficientes estandarizados nos permiten comparar la importancia relativa de diferentes atributos, ya que representan el efecto de un cambio de una desviación estándar en el atributo en la variable objetivo.
- Atributos con valores absolutos más altos de coeficientes estandarizados tienen un mayor impacto en la variable objetivo en comparación con atributos con valores absolutos más bajos.
- Podemos observar como se relacionan de manera directa con el valor de el coeficiente.

4. Error Estándar:

- El error estándar proporciona una medida de la precisión del coeficiente estimado. Cuanto menor sea el error estándar, más precisa será la estimación del coeficiente. Por ejemplo, el atributo "free.sulfur.dioxide" tiene un error estándar de aproximadamente 0.00275, lo que indica que la estimación del coeficiente para este atributo es bastante precisa.

5. Valor z:

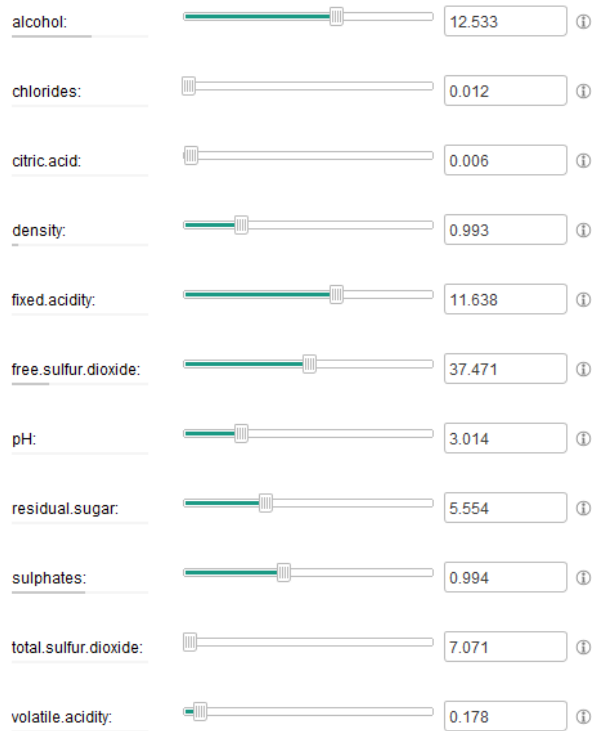
- El valor z es una medida de cuántas desviaciones estándar se encuentra el coeficiente del atributo lejos de cero. Cuanto mayor sea el valor absoluto del valor z, más lejos está el coeficiente del cero.

Análisis General:

Algunos atributos tienen coeficientes significativamente diferentes de cero y, por lo tanto, son importantes para predecir la variable objetivo. Por ejemplo, "volatile.acidity", "chlorides", "free.sulfur.dioxide", "sulphates" y "alcohol" tienen valores p muy pequeños, lo que sugiere que son factores importantes en la predicción. Sin embargo hay que tener en cuenta la significancia de cada coeficiente como "fixed.acidity", "citric.acid", "residual.sugar", "density" y "pH", tienen valores p más grandes y, por lo tanto, pueden tener una influencia menor en la variable objetivo en este modelo.

Valores del Simulador

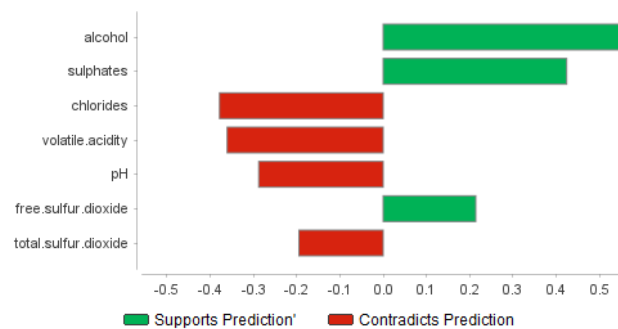
Generalized Linear Model - Simulator



Prediction

7.532

Important Factors for Prediction



Podemos observar como estos son los valores optimizados, sin embargo, aun se pueden manipular más para aumentar la predicción general, sin embargo, hay que tener en cuenta que no solo se debe buscar mejorar el resultado general, si no cuidar otras variables como la desviación estándar y la significancia de cada una de las variables utilizadas en el modelo.

Referencias

- Anáhuac Online. (2019). *Auto Model* [Archivo de video]. [Contenido creado para Anáhuac Online].
- Anáhuac Online. (2019). *Regresión lineal* [Archivo de video]. [Contenido creado para Anáhuac Online].