**UNIVERSIDAD SERGIO ARBOLEDA**

**MAESTRÍA EN MATEMÁTICAS APLICADAS**

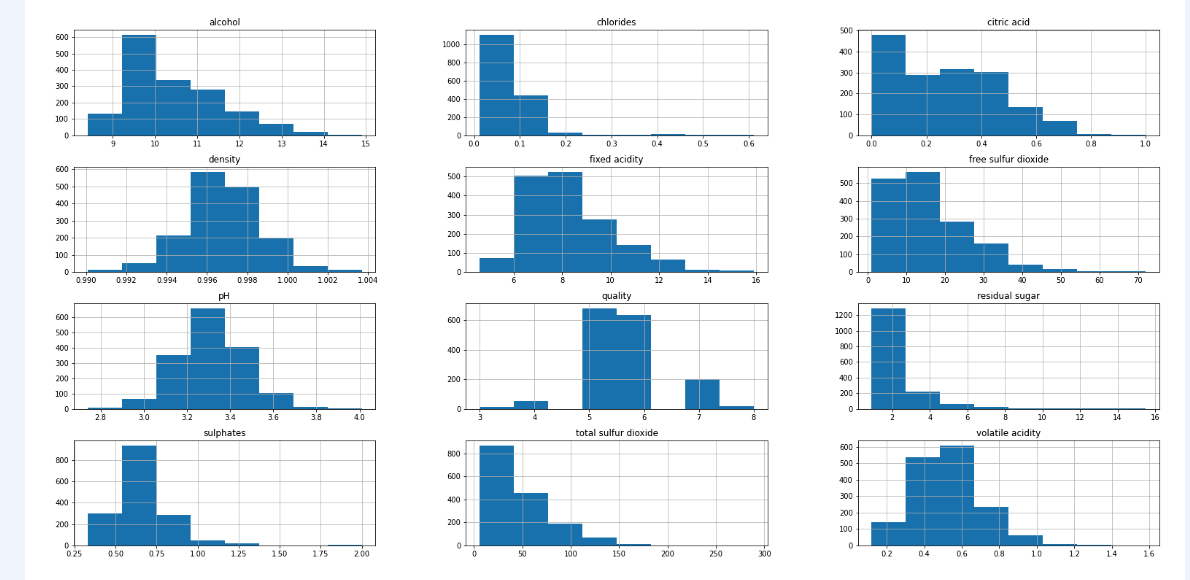
**TARABAJO FINAL**

**PROFESORA:** LUZ STELLA GÓMEZ

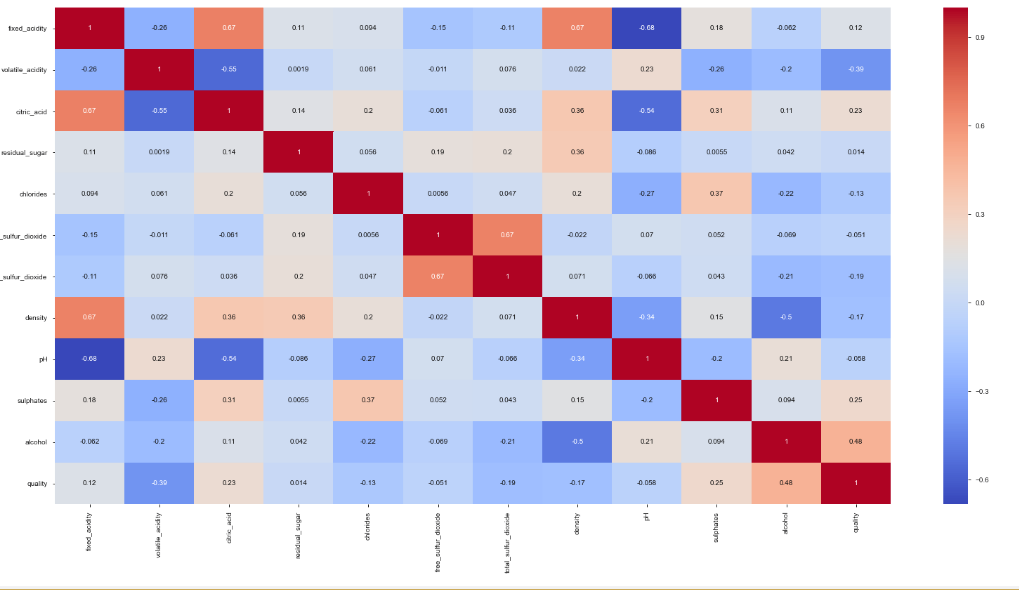
**ESTUDIANTE:** Carlos Mauricio Moreno.

El objetivo de este trabajo es usar el clasificador KNN de Python , para encontrar la calidad de un vino de acuerdo con 11 variables con las que se encuentra relacionada la misma, para emplear este clasificador se deben tener en cuenta los siguientes pasos

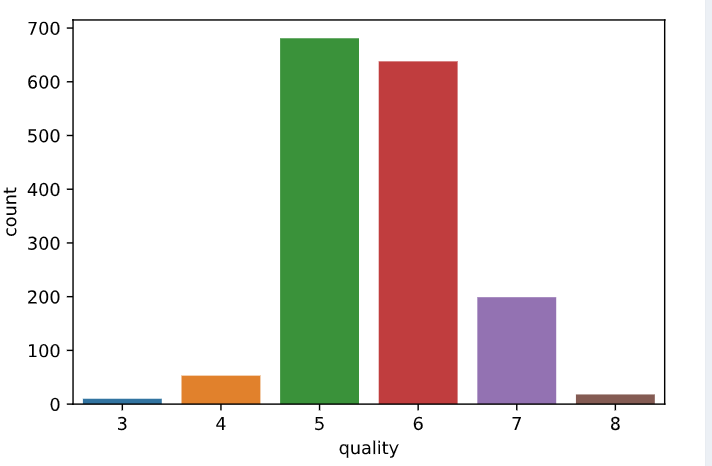
1. Se cargan los datos con lo que se va a trabajar
2. Se realizó el histograma de cada una de las variables para observar de qué forma se encuentran distribuidos los datos, de esta forma se puede observar que no se encuentran bien distribuidos en ninguna de las variables, se puede observar que la mayoría de estos se encuentran distribuidos hacia la parte izquierda de cada uno de los diagramas. Observando que se debe realizar el proceso para organizar los datos.



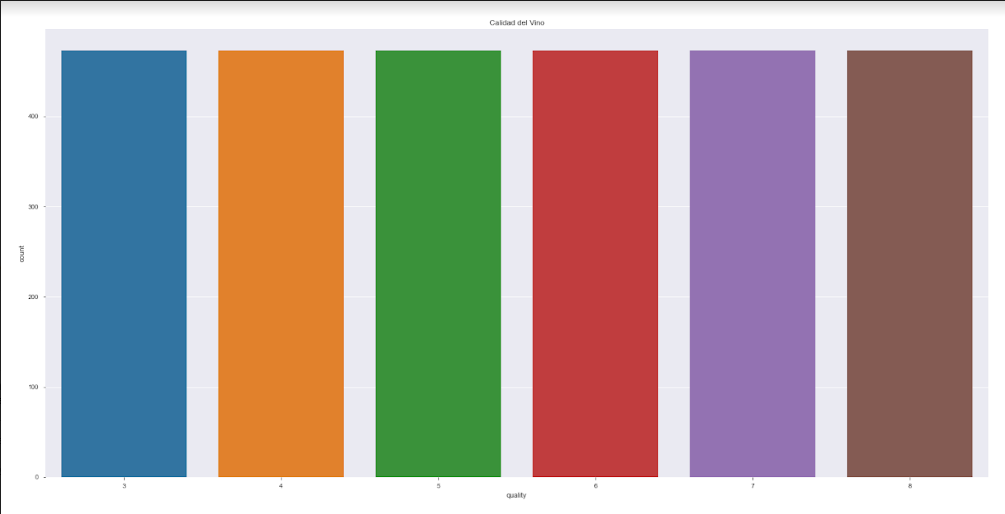
1. Para observar la correlación de los datos se realizó la matriz de correlación, obteniendo el siguiente gráfico. En el cual se puede observar que no hay ninguna correlación entre las variables.



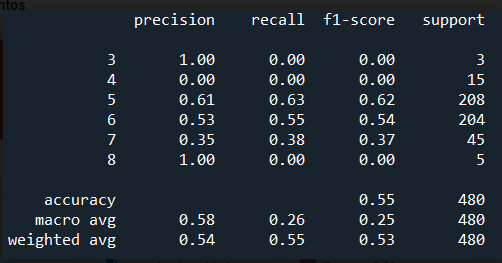
1. Se realiza una limpieza de datos observando que en ninguna de las variables existen variables nulas
2. Se realizó el diagrama de barras con la variable que se iba a relacionar con las demás variables, observando que la mayoría de vinos tiene calificaciones de 5, 6 y 7, en donde se puede observan que la mayoría de vinos se encuentran en estos rangos.



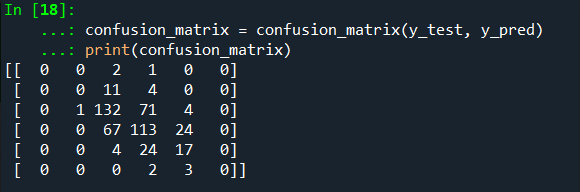
1. Se realizó la estandarización de los datos buscando que los mismos quedaran nivelados como se muestra en el gráfico.



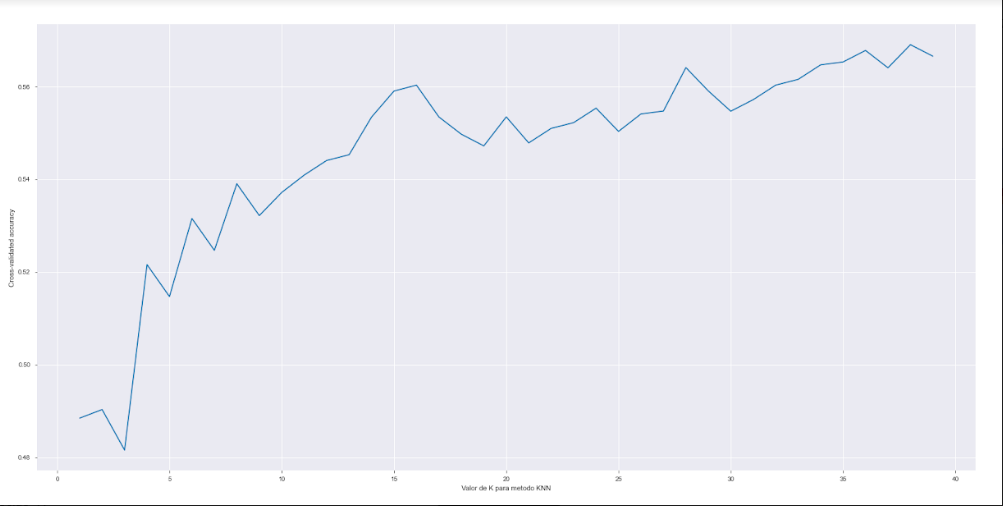
1. Se realizó un modela do de los datos mediante, lo trabajado en clase con las variables de prueba y de entrenamiento.
2. Se realizó el análisis de sensibilidad de las variables obteniendo lo siguiente, observando que algunas de las variables no muestran una gran estabilidad.



1. De acuerdo con la matriz de confusión se puede observar y comprobar lo anterior.



1. Se realizó la gráfica en la cual se puede observar el valor de Observando que uno de los valores de mayor precisión es en K=8



De acuerdo con lo anterior se puede concluir la importancia de utilizar el método de clasificación K para poder clasificar, en este caso los datos se encuentran muy difícil de clasificar los datos dados.