

# Clasificación de cervezas artesanales.

Proyecto Curso Data  
Science – 32845

Profesor: Luis Ramírez

Tutor: Anahí García

Alumno: José María Reyes  
Retana Tamayo





# Tabla de Contenidos.

1. Introducción a la temática.
2. Descripción del conjunto de datos.
3. Objetivos e hipótesis.
4. Análisis exploratorio y gráfico.
5. Algoritmos utilizados.
6. Conclusiones.



# Introducción a la temática.

- El consumo de cerveza ha existido por miles de años y ha evolucionado en términos de estilos, presentaciones y formas de consumirla. La invención de la refrigeración ha permitido la exportación de estilos locales y ha abierto mercados de una manera sin precedentes, lo que ha llevado a la creación de nuevos "sub-mercados" en los que la competencia es entre variedades con características similares que podríamos llamar "de nicho".
- Por lo tanto, es importante entender el nicho de mercado al que se quiere atender antes de producir una nueva línea de cerveza.





# Descripción del conjunto de datos.

- Se ha elegido trabajar con el data set nombrado "Beer Profiles and Ratings", obtenido de Kaggle en:  
<https://www.kaggle.com/code/aeryan/spotify-music-analysis/data>
- El siguiente estudio se centra en el análisis de un conjunto de **3197** marcas de cervezas de **934** compañías productoras de cerveza que han sido evaluadas por usuarios a través de cuestionarios en línea, así mismo, se cuenta con la información reportada por las productoras sobre sus bebidas. Las cervezas son locales y son solamente de EEUU, del mismo modo, las evaluaciones son emitidas únicamente por consumidores de dicho país.

# Descripción del conjunto de datos.

- El data set contiene **25 columnas** con información de cada una de las cervezas analizadas, de las cuales, **5 son cualitativas y 20 son cuantitativas**.
- A continuación se enlistan las variables más importantes para el estudio:
  - 🍺 ABV: contenido de alcohol (% por volumen).
  - 🍺 IBU Promedio (columna creada a través de las columnas Min IBU y Max IBU del data set): presenta el nivel de amargor de la cerveza del 0 al 100.
  - 🍺 review \_ aroma: evaluación del aroma (percepción olfativa) de la cerveza.
  - 🍺 review \_ appearance: evaluación de la apariencia (percepción visual) de la cerveza.
  - 🍺 review \_ palate: evaluación de la sensación en el paladar.
  - 🍺 review \_ taste: evaluación del sabor (percepción del gusto) de la cerveza.
  - 🍺 review \_ overall: evaluación global de la cerveza.

# Objetivos e hipótesis.

- Objetivos:

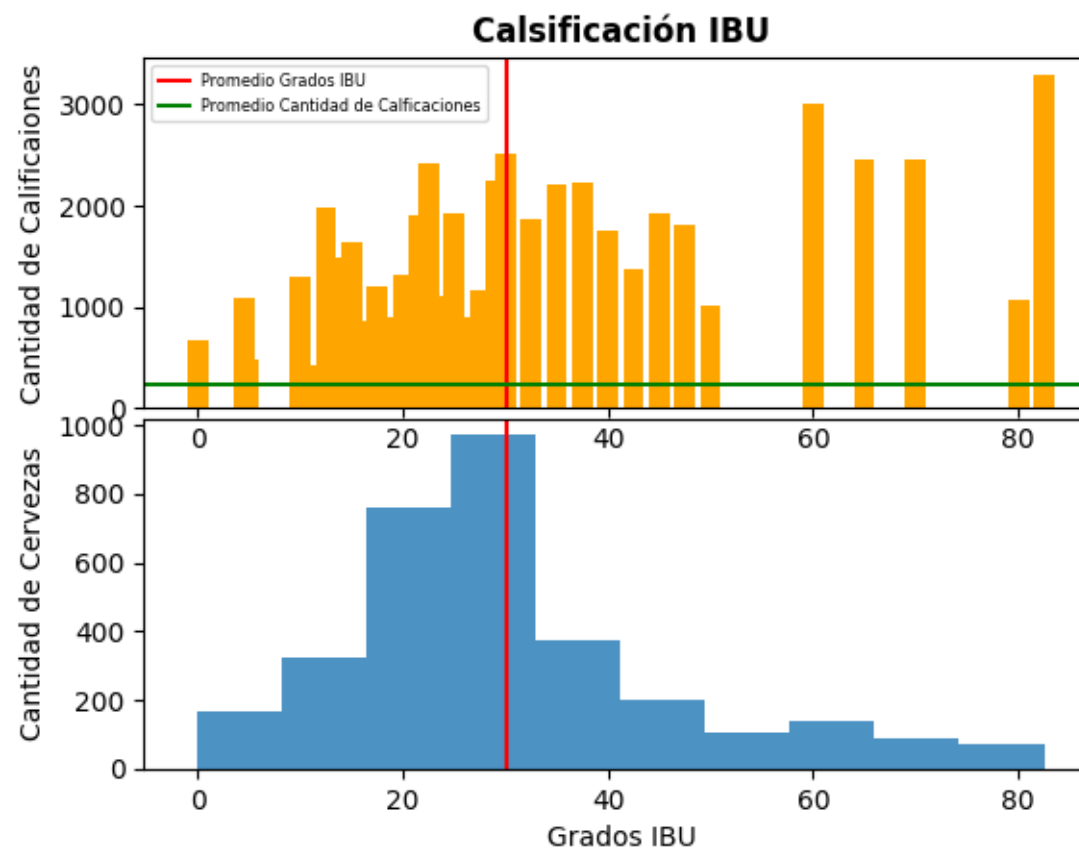
- 🍺 Encontrar las variables que más correlación tienen con la calificación global otorgada a una cerveza.
- 🍺 Poder predecir las características y variedad que producirían una cerveza muy bien calificada.

- Hipótesis:

- 1) Entre mayor sea la clasificación de grados IBU de una cerveza, esta tendrá un mercado más de nicho, por lo que: habrá menos cantidad de cervezas con grados IBU elevados.
- 2) Las cervezas con mayores grados IBU tendrán menos consumidores (por ende un menor número de calificaciones).
- 3) La calificación dada al sabor será la que mayor correlación positiva tenga con respecto a la calificación global.
- 4) Las cervezas con un porcentaje de alcohol mayor a 6° tendrán un promedio de calificación global menor que aquellas de menos de 6°.

# Análisis Exploratorio y Gráfico.

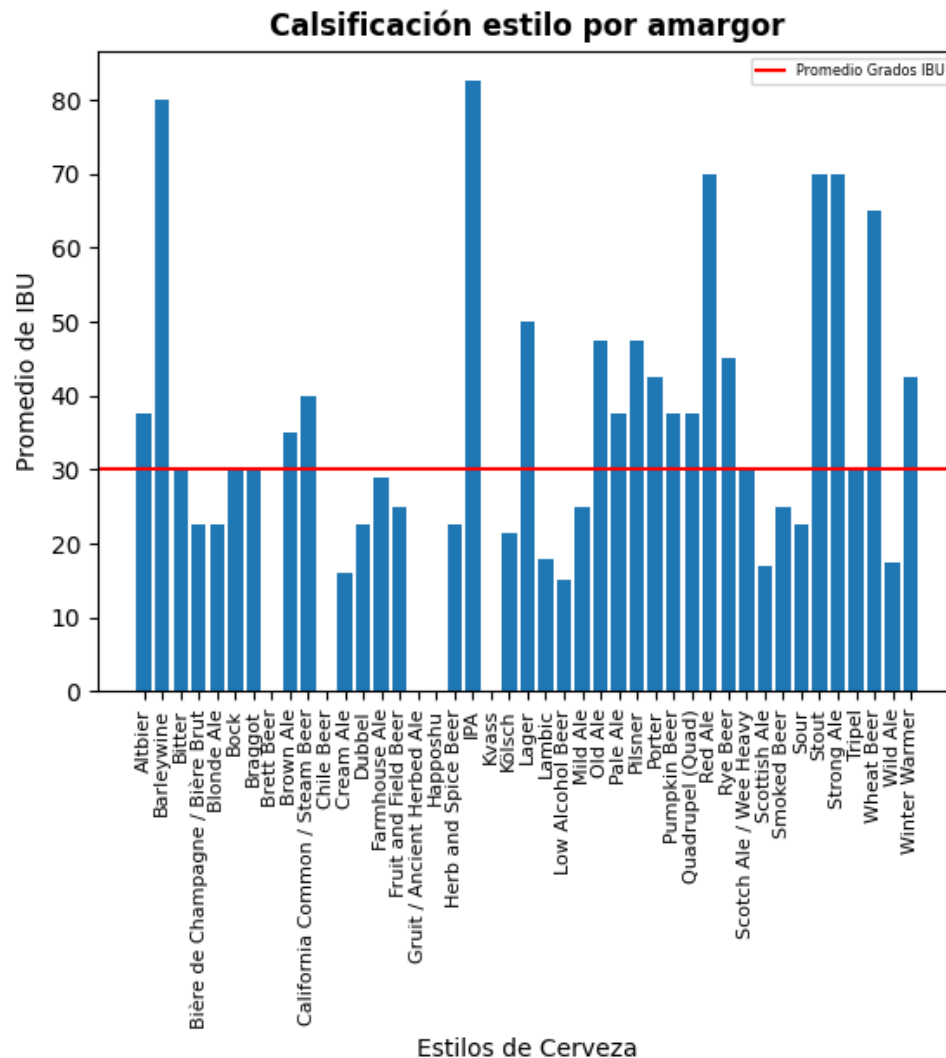
- Para la variable del IBU Promedio:



Encontramos que el promedio de graduación IBU de nuestra muestra es de  $30.08^\circ$  y que cada cerveza es calificada 233 veces en promedio.

# Análisis Exploratorio y Gráfico.

## ● IBU Promedio:



Las variedades más comunes de cerveza fueron:

Lager (567/50)

Stout (254/70)

Wheat Beer (206/65)

Las variedades con mayor graduación IBU fueron:

IPA (163/82.5)

Barleywine (64/80)

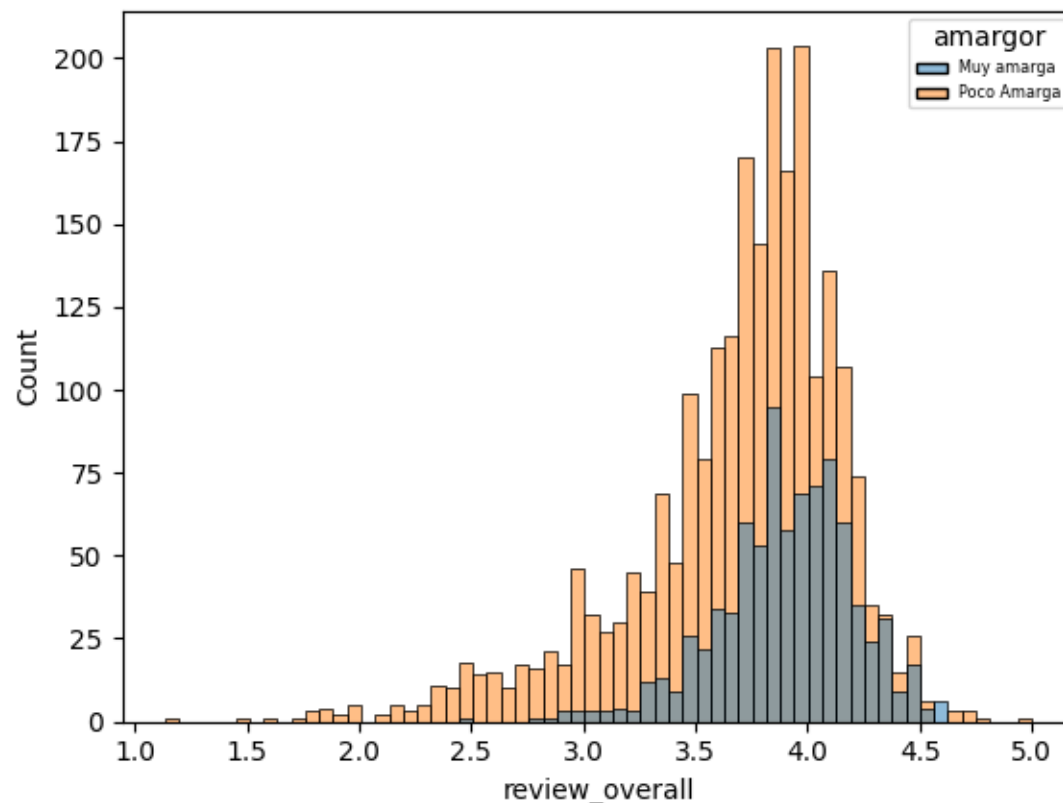
Red Ale (90/70)

(cantidad de cervezas de esta variedad/IBU Promedio máximo registrado)



# Análisis Exploratorio y Gráfico.

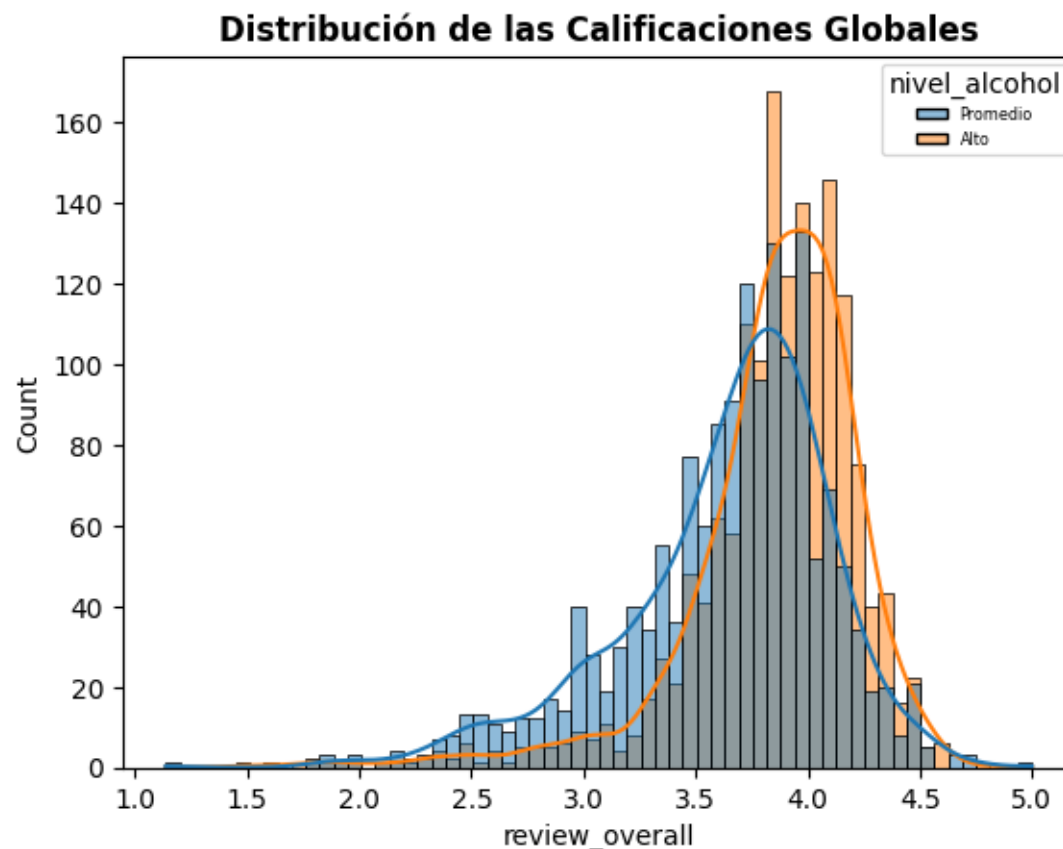
- IBU Promedio:



Podemos ver que en general, las calificaciones de las cervezas presentan una asimetría hacia la derecha, pero cuando diferenciamos entre cervezas con mayor amargor ( $IBU \geq 35$ ), la tendencia es aún más marcada.

# Análisis Exploratorio y Gráfico.

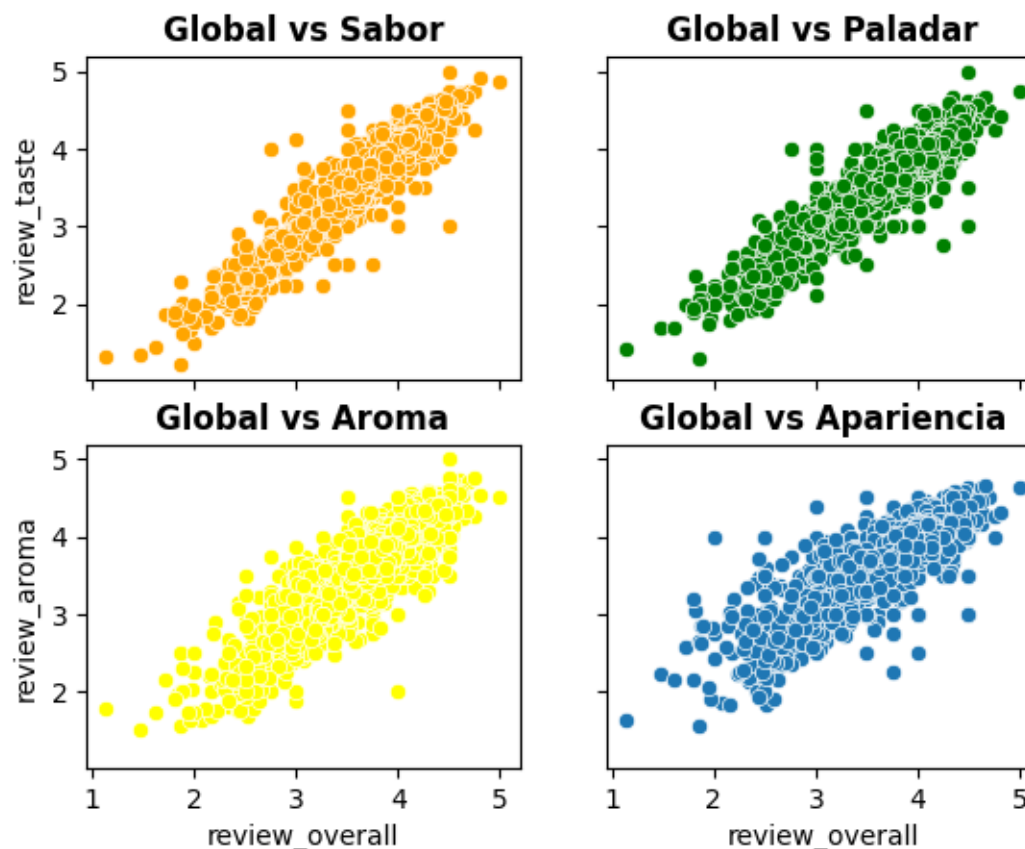
- ABV:



Similar que con los grados IBU, encontramos que a mayor graduación de alcohol ( $\geq 6^\circ$ ) habrá una asimetría hacia la derecha mayor.

# Análisis Exploratorio y Gráfico.

- Calificaciones:



Se identifica una fuerte correlación entre la calificación global y la calificación dada al sabor de una cerveza, siendo la calificación que mayor correlación guarda con la calificación global.

# Algoritmos utilizados.

- Feature selection:

🍺 Nos permitió identificar las 7 variables que más influyen sobre la variables de “calificación al sabor”, misma que ha sido otorgada por los consumidores.

- Sensación de sabores especiados (Spices)
- Sensación de amargor (Bitter)
- Graduación de alcohol (ABV)
- Sensación de mata (Malty)
- Sensación afrutada (Fruits)
- IBU Promedio.
- Salvo por ABV e IBU Promedio, todas las demás variables son subjetivas y cuyos puntajes han sido dados por los consumidores encuestados.



# Algoritmos utilizados.

- Random Forest Classification:

🍺 Se ha utilizado este algoritmo de la librería sklearn para poder entrenar a la computadora a identificar un estilo de cerveza partiendo de las variables:

🍷 ABV

🍷 IBU Promedio

🍷 Body

🍷 Astringency

🍷 Alcohol

🍷 Bitter

🍷 Sweet

🍷 Sour

🍷 Salty

🍷 Fruits

🍷 Hoppy

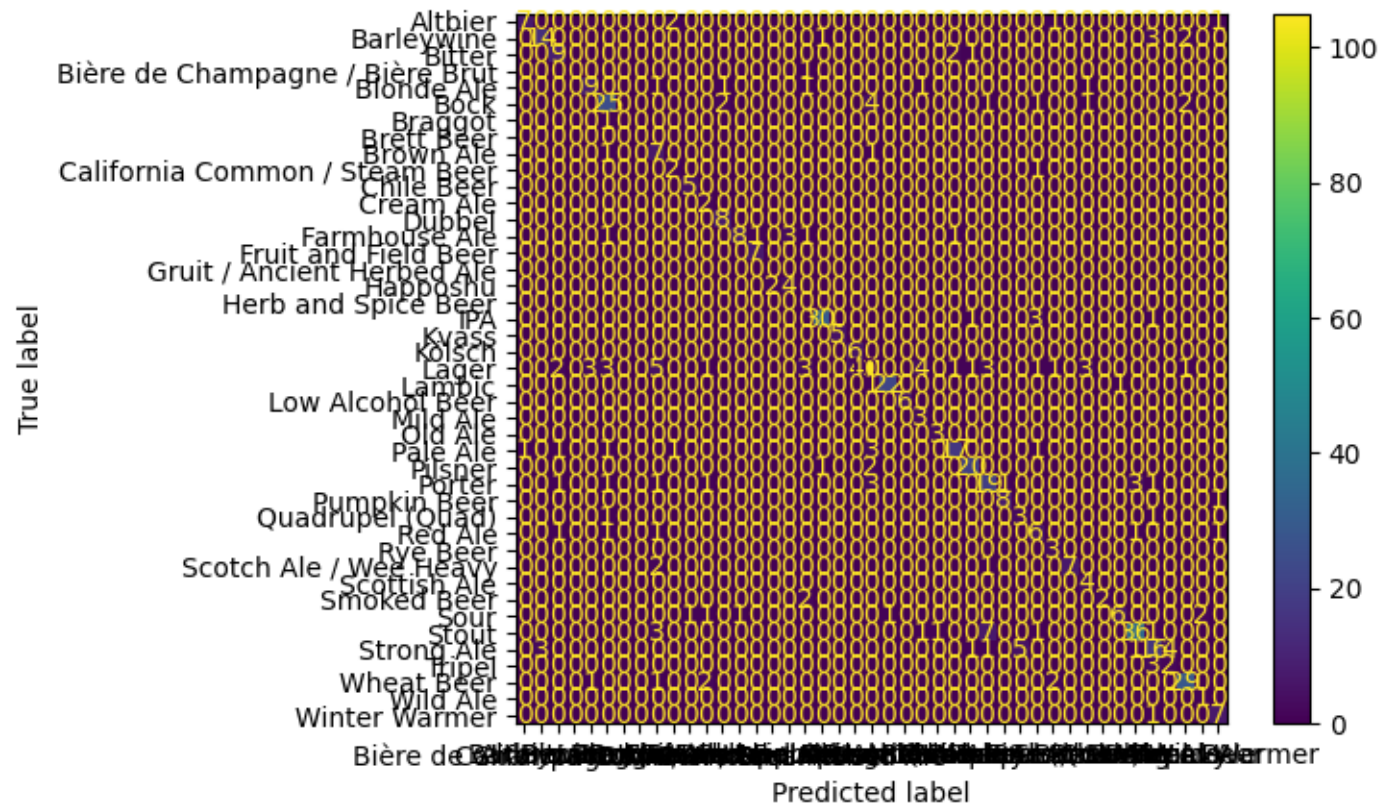
🍷 Spices

🍷 Malty

🍺 Es decir, todas las variables cuantitativas sobre la cerveza, exceptuando sus calificaciones.

# Algoritmos utilizados.

- Random Forest Classification:



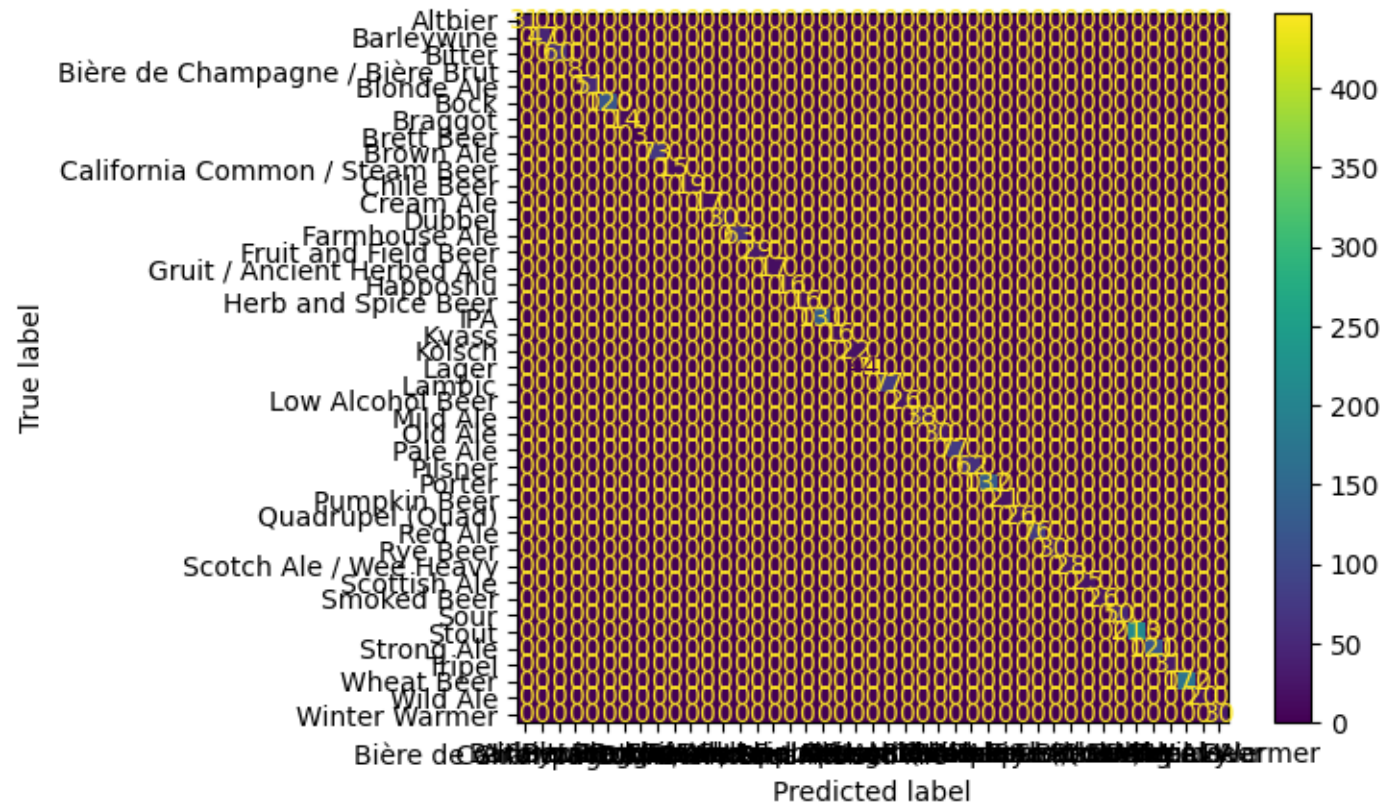
Los resultados resultan difíciles de ser visualizados pues tenemos un total de 43 variedades en el estudio; sin embargo, el modelo obtuvo las siguientes métricas promedio:

Exactitud: (66.00/94.27/98.17)  
Precisión: (0.00, 41.18, 50.00)  
Sensibilidad: (0.00/38.05/90.52)  
Especifici.: (57.40/96.28/100.00)  
F1 Score: (14.29/40.34/58.01)

(min/prom/max)

# Algoritmos utilizados.

- Random Forest Classification:



Aunque parece casi igual a la gráfica anterior, esta se obtuvo de aplicar el modelo con los datos de entrenamiento para ver si se tenía un problema de Underfitting u Overfitting, con lo que se concluye que podemos estar frente al primero.





# Conclusiones.

- Aceptamos las hipótesis 1) y 3), se rechaza la hipótesis 2) y no se puede concluir sobre la hipótesis 4) solo con estos datos.
- Se encontró que la calificación dada al sabor de una cerveza será casi igual a la calificación general que esta reciba, siendo el sabor la variable más valorada por los consumidores.
- A pesar de que las cervezas con mayor amargor ( $\geq 35^\circ$ ) no recibieron calificaciones negativas, la correlación entre IBU Promedio y las calificaciones al sabor y global fue de menos de 0.4.