

INDICES DE CONSUMO DE ALCOHOL DE CADA PAIS SEGÚN SU REPORTE DE FELICIDAD

Stefan Siegel Pacheco
Juanita Buitrago Ardila

1. Introducción

Este proyecto tiene como objetivo explorar la relación entre los índices de consumo de alcohol y los niveles de felicidad reportados en diferentes países. El consumo de alcohol es un fenómeno social y cultural que varía ampliamente entre las naciones, y se cree que puede tener un impacto en la felicidad y el bienestar de las personas.

Para abordar esta cuestión, se llevará a cabo un análisis utilizando datos obtenidos a partir de dos dataset, el primero sobre los índices de la felicidad por país en el año 2019, considerando factores como la generosidad, la percepción de corrupción y el PIB per cápita, entre otros. El segundo dataset aborda el consumo de alcohol por país, incluyendo datos como los litros de alcohol consumido y porciones de diferentes licores como la cerveza y el vino.

Con estos datos, se asignará un puntaje de felicidad a cada país y se analizará su relación con el consumo total de alcohol en esos lugares.

Así obteniendo una comprensión más de cómo el consumo de alcohol se relaciona con la felicidad en diferentes contextos culturales.

2. Desarrollo

Primero es pertinente reconocer que se hace la fusión de los dos dataset basándose en un objetivo compartido que es el país. Las características se obtienen a partir de la normalización de los datos combinados. A partir de lo anterior se implementaron tres métodos de clasificación. Para el primer método se usaron arboles de decisión, a partir del entrenamiento de este se obtienen predicciones a partir de la clasificación de los países en función de las características de los dos datasets. Por otro lado se implementa el método de regresión logística, estableciendo primeramente un umbral de alto de consumo de alcohol si es mayor a 6, a partir del entrenamiento y la aplicación del modelo utiliza las características de puntuación de la felicidad y consumo de alcohol para predecir si es alto o bajo para cada país. Por último, se implementa el método SVM con el mismo objetivo del anterior método, solo que esta vez se asocian numéricamente las palabras con 1 y 0 para alto y para bajo respectivamente

3. Resultados

Implementación de Árboles de Decisión:

- Puntuación de MCC (conjunto de entrenamiento): 0.815272474554896
- Puntuación de MCC (conjunto de prueba): 0.8249081486743332
- Puntuación F1 (conjunto de entrenamiento): 0.801169590643275

Implementación de Regresión Logística:

- Precisión del modelo: 0.9655172413793104

Implementación de SVM:

- Precisión del modelo: 0.9655172413793104
- Puntuación F1: 0.962962962962963
- Coeficiente de correlación de Matthews (MCC): 0.9330250341152236
- Puntuación F1 (conjunto de prueba): 0.8160919540229884

4. Conclusiones

- Los resultados del modelo de Árboles de Decisión muestran una buena capacidad de predicción, con altas puntuaciones de MCC tanto en el conjunto de entrenamiento como en el conjunto de prueba. Esto indica que el modelo puede clasificar con precisión las categorías.
- El modelo de Regresión Logística y el modelo SVM también mostraron un buen desempeño, con una alta precisión en la clasificación de las categorías de consumo de alcohol.
- El modelo SVM obtuvo una puntuación F1 y un coeficiente de correlación de Matthews (MCC) bastante altos, lo que indica una buena capacidad de clasificación y una correlación fuerte entre las predicciones y las etiquetas reales.
- En cuanto a los datos de felicidad y consumo de alcohol, se observó una correlación entre la puntuación de felicidad y el consumo total de alcohol puro. Sin embargo, se debe tener precaución al interpretar estos resultados, ya que el análisis se basó en datos de un solo año y no se puede establecer una relación causal definitiva.