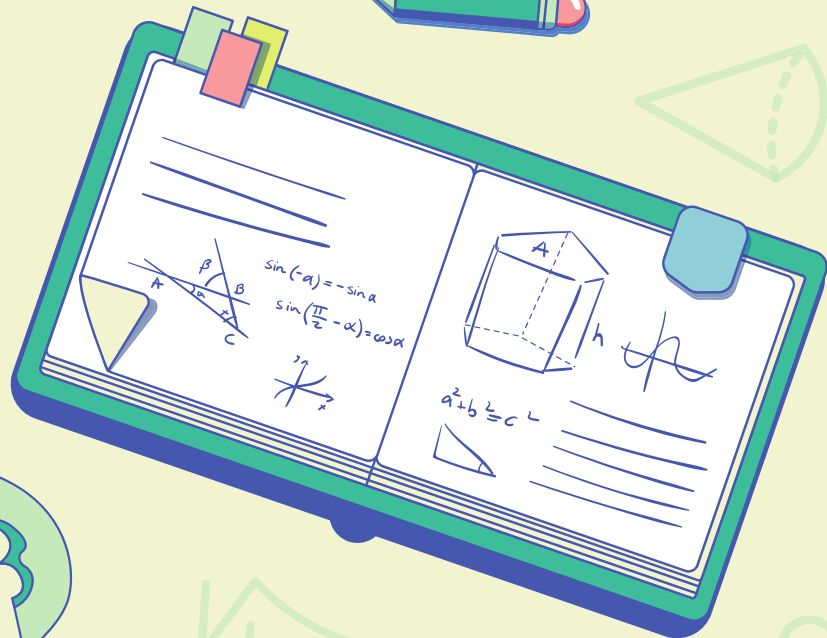


MÉTRICAS DE EVALUACIÓN

Equipo:

Daniela Paola Huerta Torres
Verónica Vanessa Aguilar Ortiz
Fernando Jahir Rodríguez Chaires
Miguel Ángel Jalomo Hernández

1872932
1855188
1751125
1857876





Métricas de evaluación

Se utilizan para valorar el rendimiento de un modelo de aprendizaje automático. Tiene como objetivo estimar la precisión de la generalización de un modelo sobre los datos futuros (no vistos/fuera de muestra).

Matriz de Confusión.

		Valores Predichos	
		Negativo	Positivo
Valores Actuales	Negativo	Verdadero Negativo	Falso Positivo (Error Tipo I)
	Positivo	Falso Negativo (Error Tipo II)	Verdadero Positivo

Tipos de Errores

Error de Tipo I: equivalente a los Falsos Positivos

El primer tipo de error posible implica el rechazo de una hipótesis nula que es verdadera.

Error de tipo II: equivalente a Falsos Negativos

El otro tipo de error que ocurre cuando aceptamos una hipótesis falsa nula. Este tipo de error se llama error de tipo II y también se conoce como error de segundo tipo.

Tipos de Errores

Error Tipo I (Falsos Positivos)

Afirmar que un medicamento tiene un efecto positivo contra una enfermedad cuando en realidad no lo tiene.



Error Tipo II (Falsos Negativos)

Afirmar que un medicamento no tiene un efecto positivo contra una enfermedad cuando en realidad si lo tiene.



Tasas

Exactitud

Es el cociente entre las predicciones correctas y la proporción de todas las predicciones hechas.

$$\frac{VN + VP}{Total}$$

$$Total = VN + FN + FP + VP$$

Tasa de error

Frecuencia de error. Cociente entre predicciones incorrectas y la proporción de todas las predicciones hechas.

$$\frac{FN + FP}{Total}$$

$$Total = VN + FN + FP + VP$$





Precisión

Predice con que frecuencia es correcto. Cociente entre la frecuencia de los valores uno correctos y el total de valores uno observados.

$$\frac{VP}{FP + VP}$$

Especificidad

Se calcula como el numero de predicciones negativas correctas dividido entre el numero total de negativos.

$$\frac{VN}{VN + FP}$$



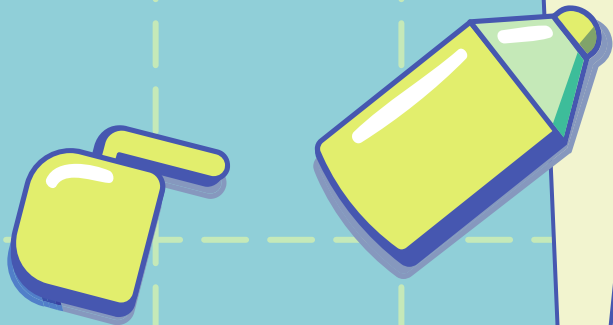
Exhaustividad

Predice con que frecuencia se predice un valor verdadero cuando en realidad si es verdadero. Cociente entre los valores uno corrector y el total de valores uno observado.

$$\frac{VP}{VP + FN}$$



Puntaje F1



El puntaje F1 es el promedio ponderado de precisión y sensibilidad. Por lo tanto, esta puntuación tiene en cuenta tanto los falsos positivos como los falsos negativos.

$$\text{Puntaje F1} = 2 * \text{Precisión} * \text{Sensibilidad} / (\text{Precisión} + \text{Sensibilidad})$$

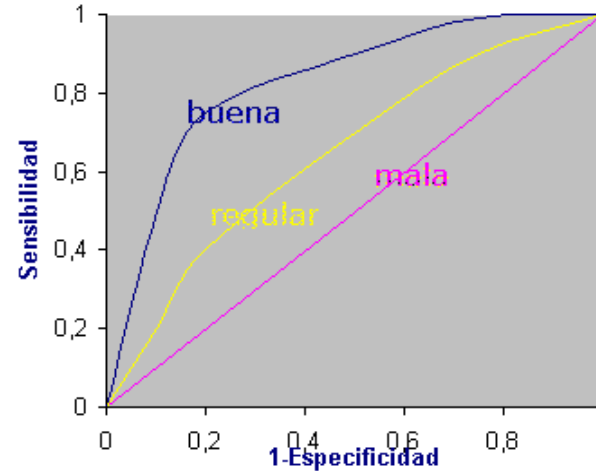
Este podría ser una métrica de evaluación efectiva en los siguientes escenarios de clasificación:

- Cuando los Falsos Positivos y la Falsos Negativos son igualmente costosos.
- Añadir más datos no cambia el resultado de manera efectiva.
- El Verdadero Negativo es alto (como en las predicciones de inundaciones, predicciones de cáncer, etc.)

Curva de características operativas del receptor (ROC)

La curva ROC es una representación gráfica del rendimiento del clasificador que muestra la distribución de las fracciones de verdaderos positivos y de falsos positivos. La fracción de verdaderos positivos se conoce como sensibilidad, sería la probabilidad de clasificar correctamente a un individuo cuyo estado real sea definido como positivo.

Tipos de curvas ROC

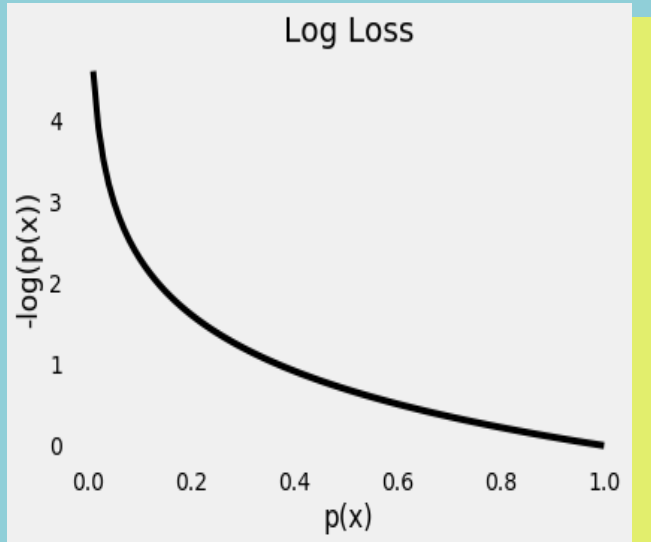


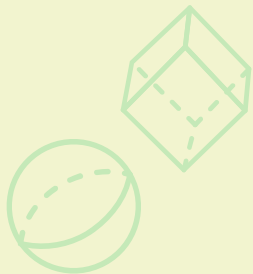


Pérdida logarítmica

La pérdida logarítmica es la métrica de clasificación más importante basada en probabilidades.

Mide el desempeño de un modelo de clasificación en el que la entrada de la predicción es un valor de probabilidad entre 0 y 1. Una pérdida logarítmica menor es mejor, con un modelo perfecto teniendo una pérdida logarítmica de 0.





Índice Jaccard

El índice Jaccard es una de las formas más simples de calcular y averiguar la exactitud de un modelo de clasificación de aprendizaje automático.

El índice Jaccard es una estadística utilizada para comprender las similitudes entre los conjuntos de muestras. La medición enfatiza la similitud entre conjuntos de muestras finitas y se define formalmente como el tamaño de la intersección dividido por el tamaño de la unión de los dos conjuntos etiquetados, con la fórmula como:

$$J(y, y1) = \frac{|y \cap y1|}{|y \cup y1|} = \frac{|y \cap y1|}{|y| + |y1| - |y \cap y1|}$$

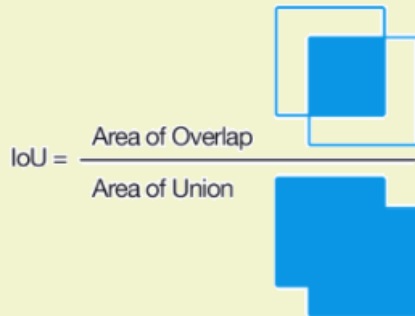
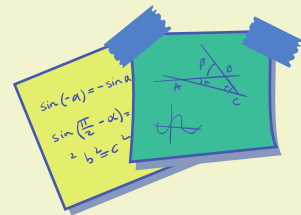
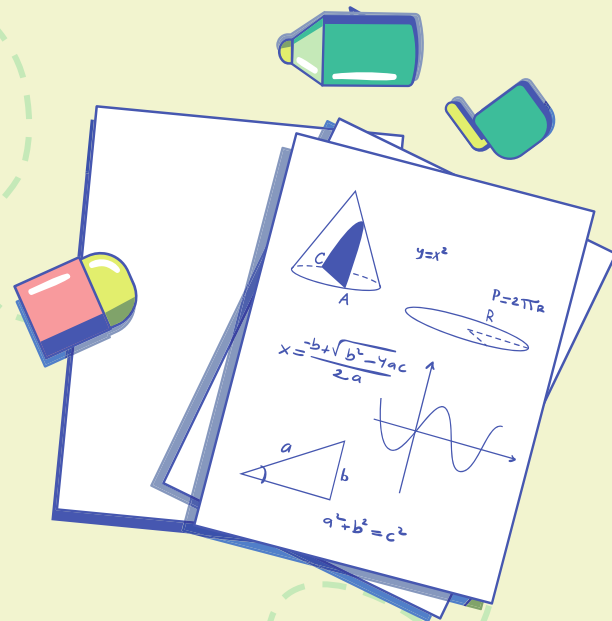
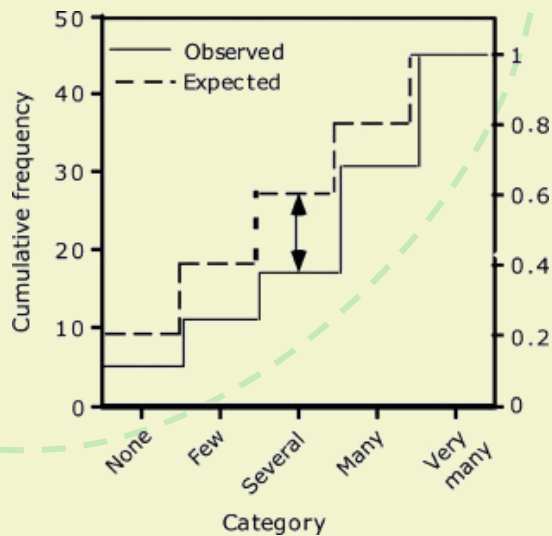


Gráfico de Kolmogorov Smirnov

El gráfico K-S o Kolmogorov-Smirnov mide el rendimiento de los modelos de clasificación. Más exactamente, K-S es una medida del grado de separación entre las distribuciones positivas y negativas.



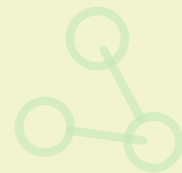
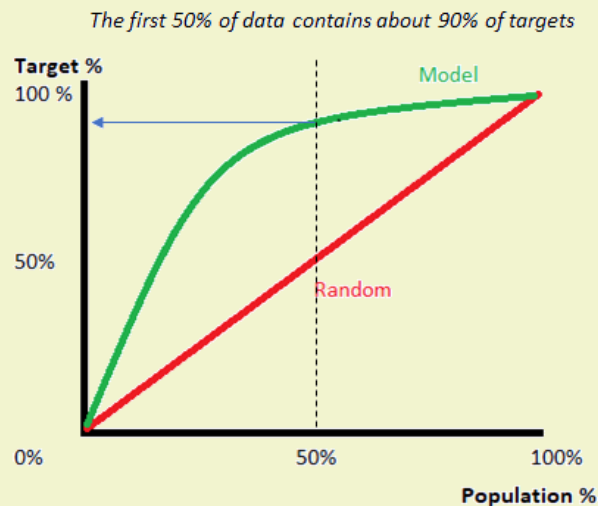
$$\frac{a \times b}{x}$$

En la mayoría de los modelos de clasificación la K-S caerá entre 0 y 100, y cuanto más alto sea el valor mejor será el modelo para separar los casos positivos de los negativos.



Gráfico de ganancia y elevación

La ganancia o el levantamiento es una medida de la eficacia de un modelo de clasificación calculado como la relación entre los resultados obtenidos con y sin el modelo. Los gráficos de ganancia y elevación son ayudas visuales para evaluar el rendimiento de los modelos de clasificación. Cuanto mayor sea la elevación (es decir, cuanto más lejos esté de la línea de base), mejor será el modelo.



Coeficiente de Gini $x-y$

El coeficiente de Gini es una métrica popular para los valores de clase desequilibrados. El coeficiente oscila entre 0 y 1, donde 0 representa la igualdad perfecta y 1 la desigualdad perfecta. Aquí, si el valor de un índice es mayor, entonces los datos estarán más dispersos.

$$(x-y)^2$$

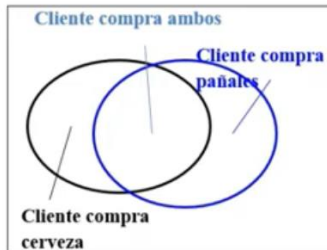


Asociación

Las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

"Antecedente => Consecuente {Soporte, Confianza}"

- **Soporte** (s): probabilidad de que una transacción contenga {X & Y & Z}
- **Confianza** (c): probabilidad condicional $P(Z|X \& Y)$



Base de datos con 4 ítems y 5 transacciones

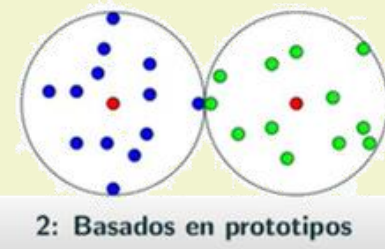
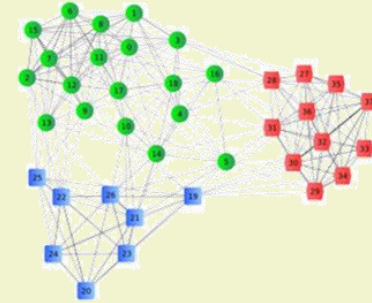
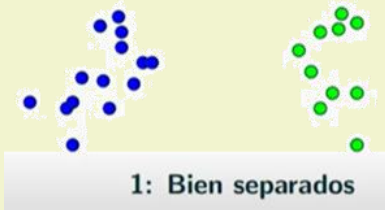
ID	Leche	Pan	Mantequilla	Cerveza
1	1	1	0	0
2	0	1	1	0
3	0	0	0	1
4	1	1	1	0
5	0	1	0	0

$I = \{\text{Leche, Pan, Mantequilla, Cerveza}\}$

Agrupamiento

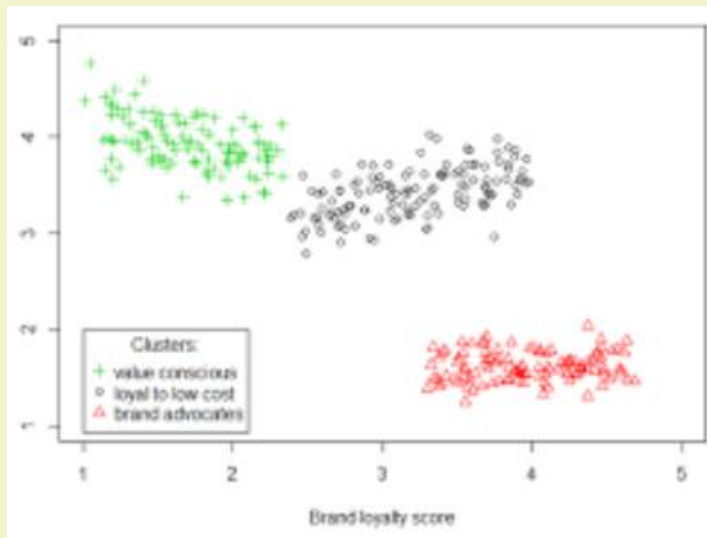
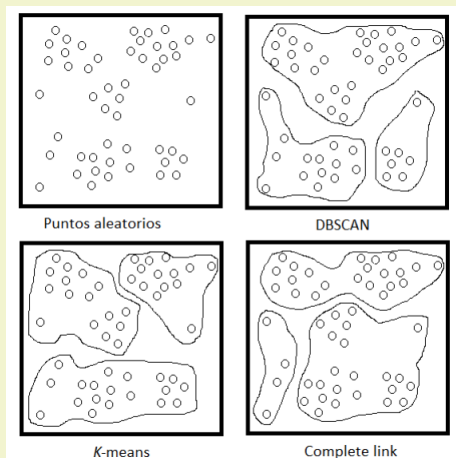


El Clustering (Agrupamiento) es un proceso no supervisado en Minería de Datos y en el Reconocimiento de Patrones, ampliamente utilizado y que es especialmente sensible a los parámetros de la entrada.



Verificar la agrupación final

Es de importancia evaluar el resultado de los algoritmos de clustering, sin embargo, es difícil definir cuando el resultado de un agrupamiento es aceptable. Por esta razón existen técnicas e índices para la validación de un agrupamiento realizado.



Gracias por su Atención

Fuentes de consulta:

<https://bookdown.org/content/2274/metodos-de-clasificacion.html#validacion-cruzada>

<https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico>

<https://aprendeia.com/evaluando-el-error-en-los-modelos-de-clasificacion-machine-learning/>

<https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/>

