# Aprendizaje Supervisado: Clasificación

TAE IA
Formación de Talento Altamente Especializado

# Aprendizaje Supervisado: Clasificación

## Definición

**Clasificación:** Cuando la variable de respuesta del modelo es categórica.

## Tipos de clasificación

- **Binaria:** Sólo se utilizan dos categorías.
- **Multiclase**: más de dos  etiquetas de clase.

TAE IA
Formación de Talento Altamente Especializado

# Algoritmos de Clasificación

- Redes neuronales
- Algoritmos basados en árboles
- Máquinas de soporte vectorial
- Regresión Logística
- Naive Bayes
- K-Vecinos más cercanos

TAE IA
Formación de Talento Altamente Especializado

# Algoritmos de Clasificación

- Redes neuronales
- Algoritmos basados en árboles
- Máquinas de soporte vectorial
- Regresión Logística
- Naive Bayes
- K-Vecinos más cercanos

TAE IA
Formación de Talento Altamente Especializado

# Algoritmos de Clasificación

Regresión Logística

- Simplicidad
- Velocidad
- Entendimiento

TAE IA
Formación de Talento Altamente Especializado

# Regresión Logística

## Definición

**Regresión Logística:** Modelo estadístico utilizado para predecir un coeficiente asociado a la probabilidad de una variable binaria.

## Detalles

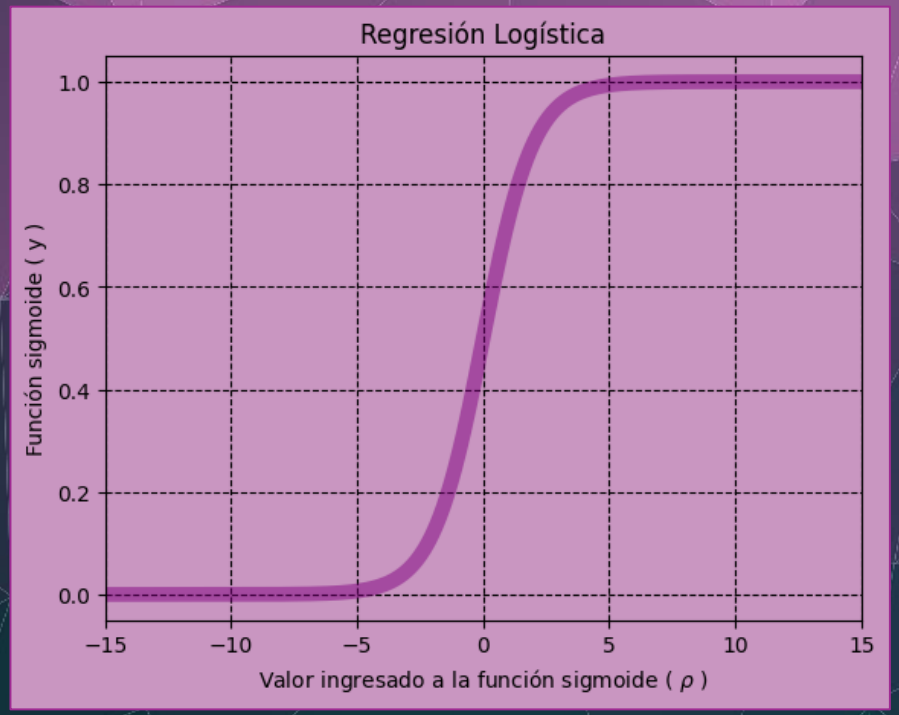
Como ya veremos difiere de la regresión lineal que predice valores continuos, en este caso nos limitamos a clasificar en categorías denotadas usualmente como 1 (presencia de la clase) y 0 (ausencia de la clase).

TAE IA
Formación de Talento Altamente Especializado

# Regresión Logística

Función Sigmoide

$$y = \frac{1}{1 + e^{-\rho}}$$

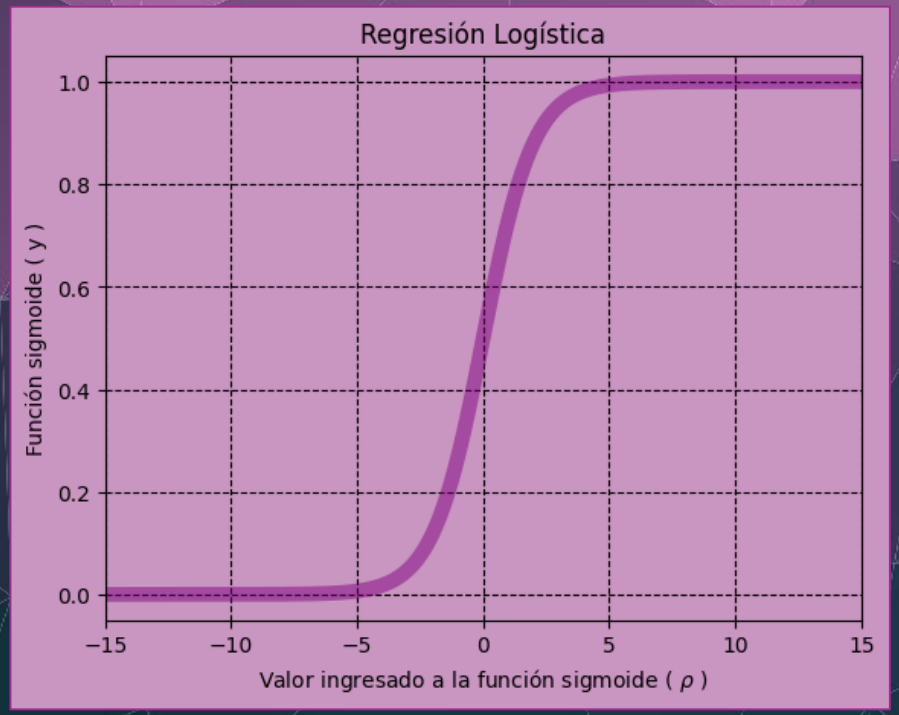



TAE IA
Formación de Talento Altamente Especializado

# Regresión Logística

Función Sigmoide

$$y = \frac{1}{1 + e^{-\rho}}$$




$$\rho = a_0 + a_1 X_1 + a_2 X_2 + \ \ldots + a_n X_n$$

TAE IA
Formación de Talento Altamente Especializado

# Regresión Logística

$$\rho = a_0 + a_1X_1 + a_2X_2 + \ ... + a_nX_n$$

La solución consiste en encontrar el coeficiente $a_i$, asociado a la variable $X_i$. Lo encontraremos este para cada feature que tengamos disponible.

Se calcula las multiplicaciones y sumas. Así obtenemos el valor $\rho$ que se lo pasaremos a la función sigmoide, y nos entregará un valor contínuo entre 0 y 1.

TAE IA
Formación de Talento Altamente Especializado

# Regresión Logística

¡Vamos a probar!

TAE IA
Formación de Talento Altamente Especializado

# Métricas de Clasificación

TAE IA
Formación de Talento Altamente Especializado

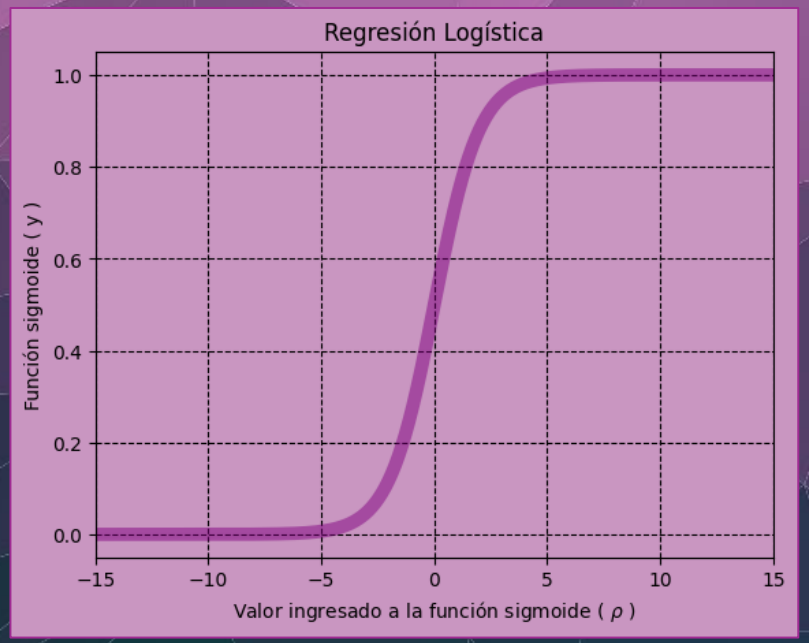
# Métricas de Clasificación

Umbral de clasificación /
Punto de corte

TAE IA
Formación de Talento Altamente Especializado

# Métricas de Clasificación

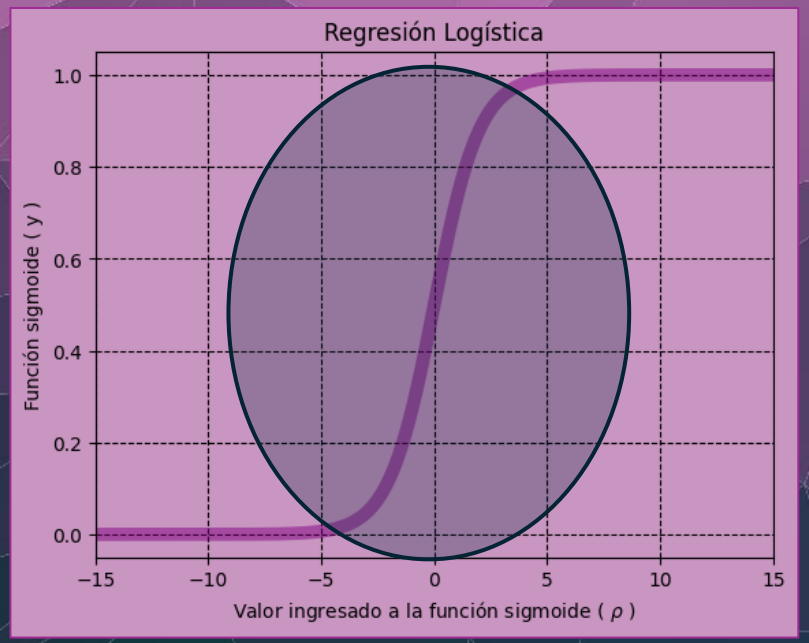
Umbral de clasificación /
Punto de corte




TAE IA
Formación de Talento Altamente Especializado

# Métricas de Clasificación

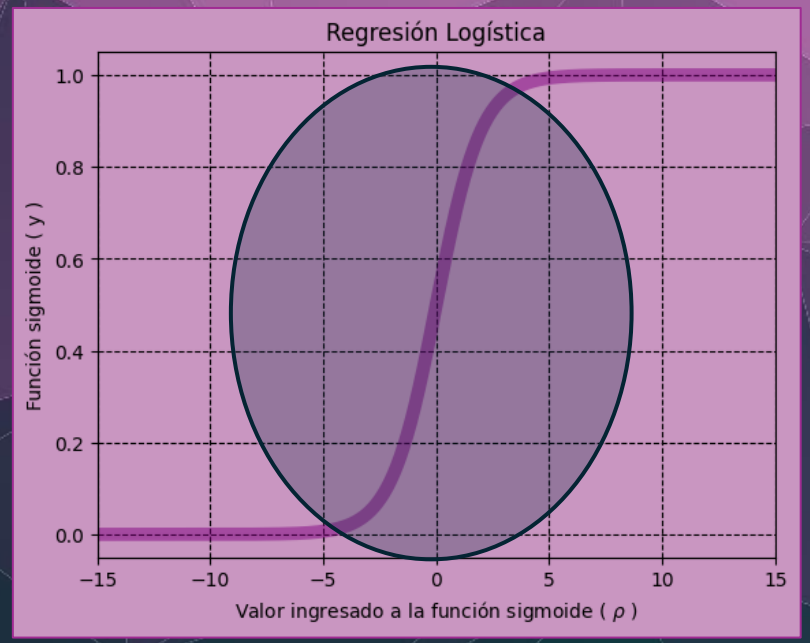
Umbral de clasificación /
Punto de corte

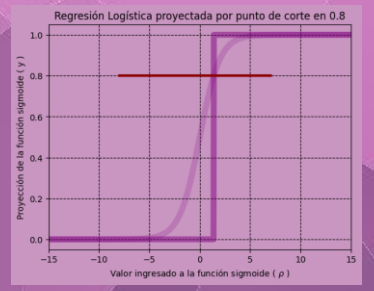


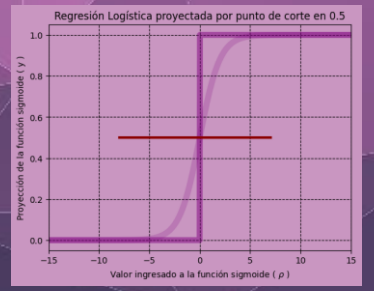

TAE IA
Formación de Talento Altamente Especializado

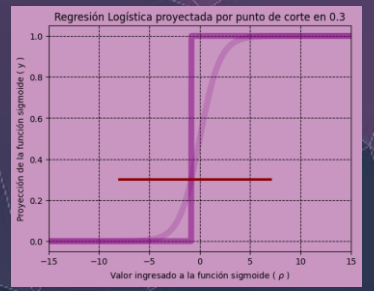
# Métricas de Clasificación

Umbral de clasificación /
Punto de corte



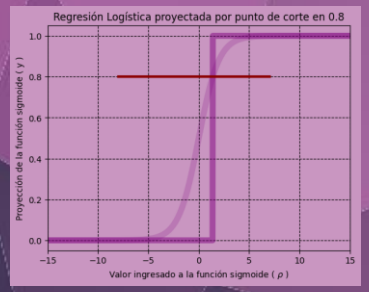



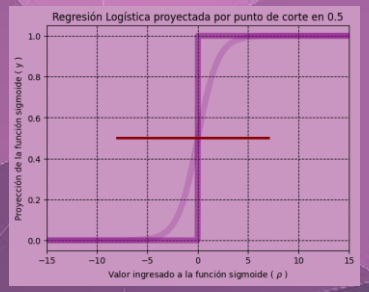



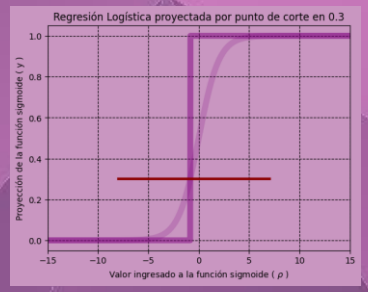




TAE IA
Formación de Talento Altamente Especializado

# Métricas de Clasificación




- Reglas de Negocio
- Reglas estadísticas (Índice de Youden)
- Definición arbitraria (0.5 en predict de modelos)
- Todo esto va a afectar las métricas del modelo.

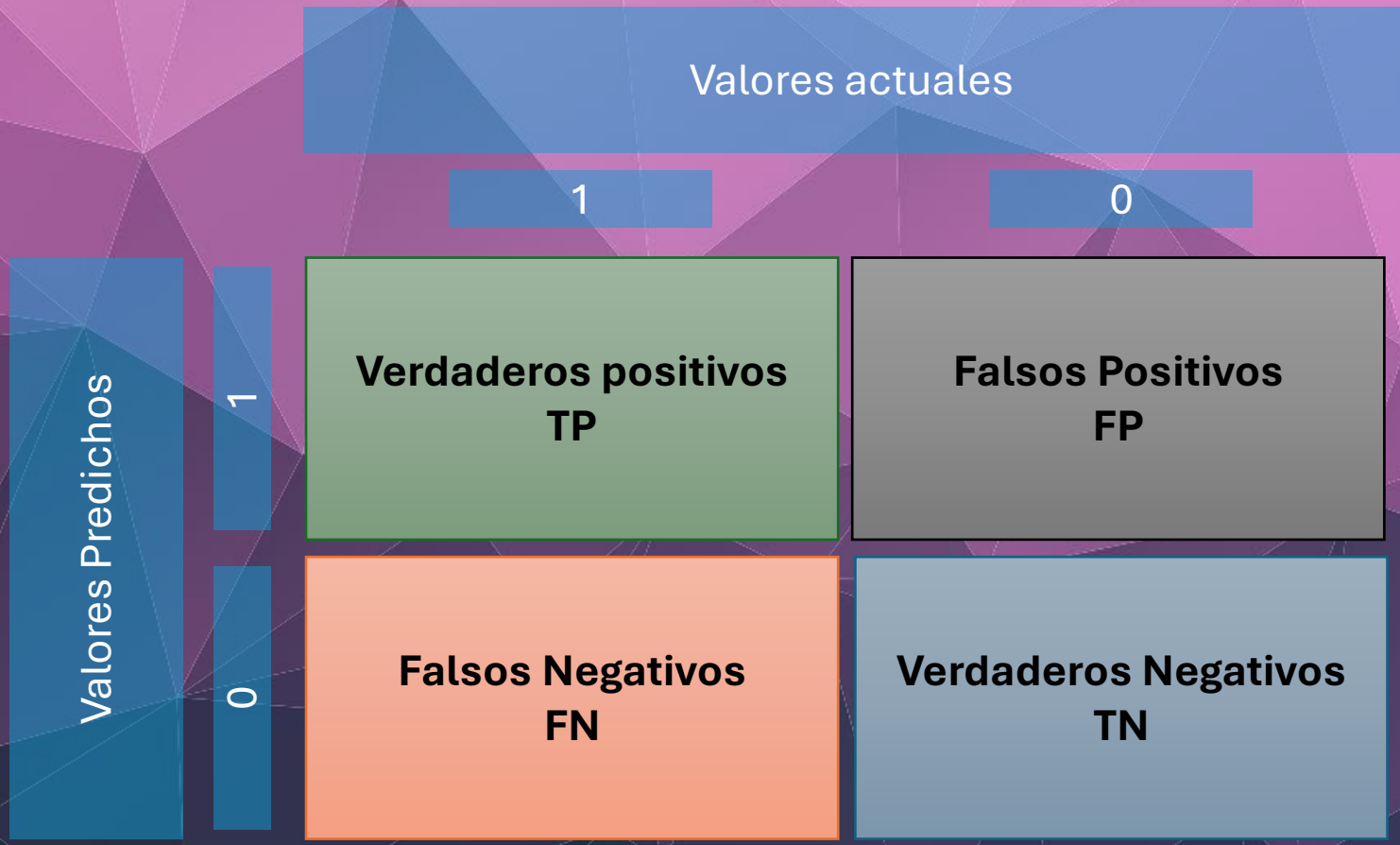
TAE IA
Formación de Talento Altamente Especializado

# Métricas de Clasificación

- Umbral no definido
  - AUC ROC
  - AUC PR
- Umbral definido
  - Matriz de confusión
    - Exactitud
    - Precisión
    - Sensibilidad
    - Coeficiente F-1

TAE IA
Formación de Talento Altamente Especializado

# Matriz de confusión




TAE IA
Formación de Talento Altamente Especializado

# Matriz de confusión

| | | |
|---|---|---|
| Exactitud (Accuracy) | $\frac{TP + TN}{TP + TN + FP + FN}$ | De todos los casos, ¿a cuántos les atiné? |
| Precisión (Precision) | $\frac{TP}{TP + FP}$ | De todos los que predije como 1, ¿a cuantos les atiné? |
| Sensibilidad (Recall) | $\frac{TP}{TP + FN}$ | De todos los que en realidad eran 1, ¿cuántos predije correctamente? |
| F1-score | $2\frac{P \; R}{P + R}$ | Promedio armónico entre Precisión y Sensibilidad |

TAE IA
Formación de Talento Altamente Especializado

# AUC-ROC

Tasa de verdaderos positivos (Recall)

$$\frac{TP}{TP + FN}$$

De todos los que en realidad eran 1, ¿cuántos predije correctamente?

Tasa de falsos positivos

$$\frac{FP}{FP + TN}$$

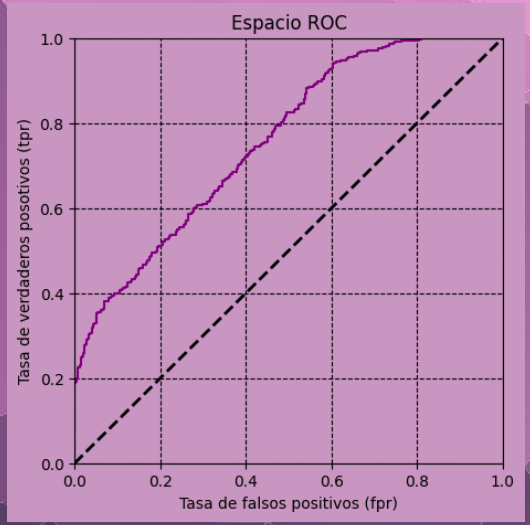
De todos los que en realidad eran 0, ¿cuantos predije erroneamente?

TAE IA
Formación de Talento Altamente Especializado

# AUC-ROC

Tasa de verdaderos positivos (Recall)

Tasa de falsos positivos




De todos los que en realidad eran 1, ¿cuántos predije correctamente?

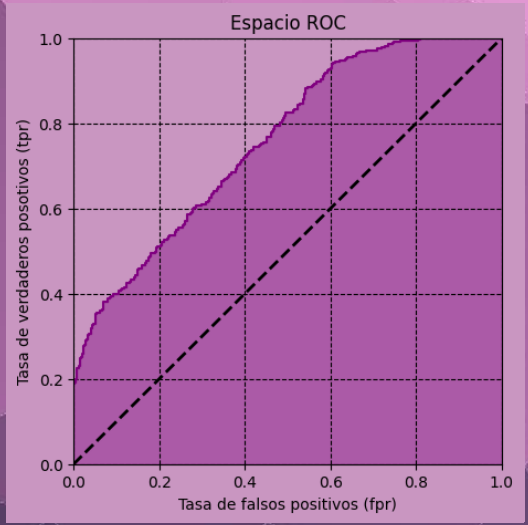
De todos los que en realidad eran 0, ¿cuantos predije erroneamente?

TAE IA
Formación de Talento Altamente Especializado

# AUC-ROC

Tasa de verdaderos positivos (Recall)

Tasa de falsos positivos

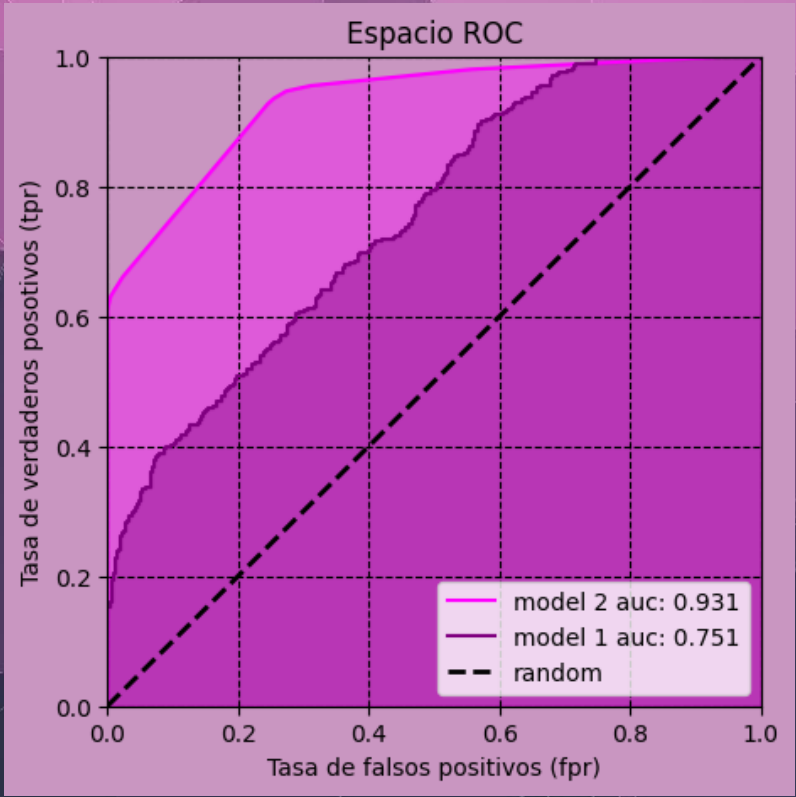



De todos los que en realidad eran 1, ¿cuántos predije correctamente?

De todos los que en realidad eran 0, ¿cuantos predije erroneamente?

TAE IA
Formación de Talento Altamente Especializado

# AUC-ROC




Podemos basarnos en esto como ayuda:

- < 0.5 Modelo es pésimo, mejor no usarlo.
- = 0.5 Modelo es igual que lanzar una moneda.
- (0.5, 0.6) Modelo es malo, pero mejor que la suerte.
- [0.6, 0.75) Modelo es regular (Zona usual).
- [0.75, 0.80) Modelo es aceptable (Zona usual).
- [0.80, 0.90) Modelo es bueno (Zona Feliz).
- [0.90, 0.96) Modelo es muy bueno (Cuidado).
- [0.96, 1.0) Modelo es excelente (Cuidado).

TAE IA
Formación de Talento Altamente Especializado

# AUC-ROC

¡Vamos a probar!

TAE IA
Formación de Talento Altamente Especializado

# Referencias


- Nusinovici, S.; Tham, Y.; Yan, M.; Ting, D.; Li, J.; Sabanayagam, C.; Wong, T.; y Cheng, C. (2020) Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of Clinical Epidemiology 122. pp. 56-69.
- Fernandes, A.; Filho, D.; Rocha, E.; Nascimento, W. (2020) Read this paper if you want to learn logistic regression. Revista de Sociologia e Política. DOI 10.1590/1678-987320287406en
- Sen, P.; Hajra, M. y Ghosh, M. (2020) Supervised Classification Algorithms in Machine Learning: A Survey and Review. Springer Nature Singapore Pte Ltd. Emerging Technology in Modelling and Graphics, Advances in Intelligent Systems and Computing 937, https://doi.org/10.1007/978-981-13-7403-6_11
- AWS (2023) ¿Qué es la regresión logística? Recuperado el 14 de julio del 2024 de https://aws.amazon.com/es/what-is/logistic-regression/
- IBM (2023) Regresión Logística. Recuperado el 14 de julio del 2024 de https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-logistic
- Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. doi:10.1016/j.patrec.2005.10.010
- Hastie, T.; Tibshirani, R. and Friedman, J. (2008) The elements of statistical learning Data Mining, Inference, and Prediction. Springer. Standford California.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.


TAE IA
Formación de Talento Altamente Especializado

# Gracias por su atención

# ¿Dudas?