

# Prediccion de Ataque Cerebral

Gabriel Velez  
Joaquin Orbe  
Gabriel Torres

# Agenda

- **Introducción**
- **Correcciones**
- **Algoritmo y mejor tecnica de entrenamiento**
- **Metricas de evaluación y curvas ROC**
- **Costo de Implementación**
- **Enlace GitHub**
- **Conclusiones**
- **Referencias**

# Introducción

- Es necesario primero manipular la base de datos para luego poder trabajar con esta y encontrar las variables que más influyen en la variable de respuesta, probar distintos modelos y finalmente escoger el mejor.
- Objetivo: Determinar si una persona tiene una alta o baja probabilidad de sufrir un ataque cerebral.
- Lo que más nos ayuda, es saber si el paciente sufrió un ataque cerebral o no.

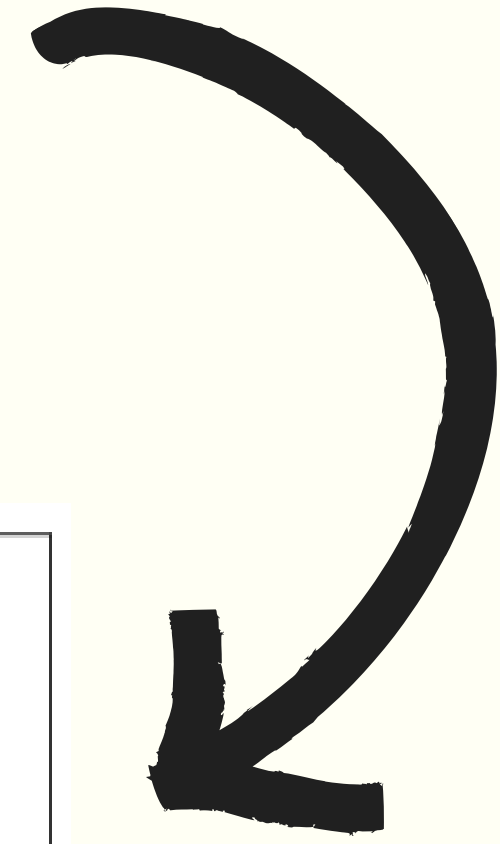
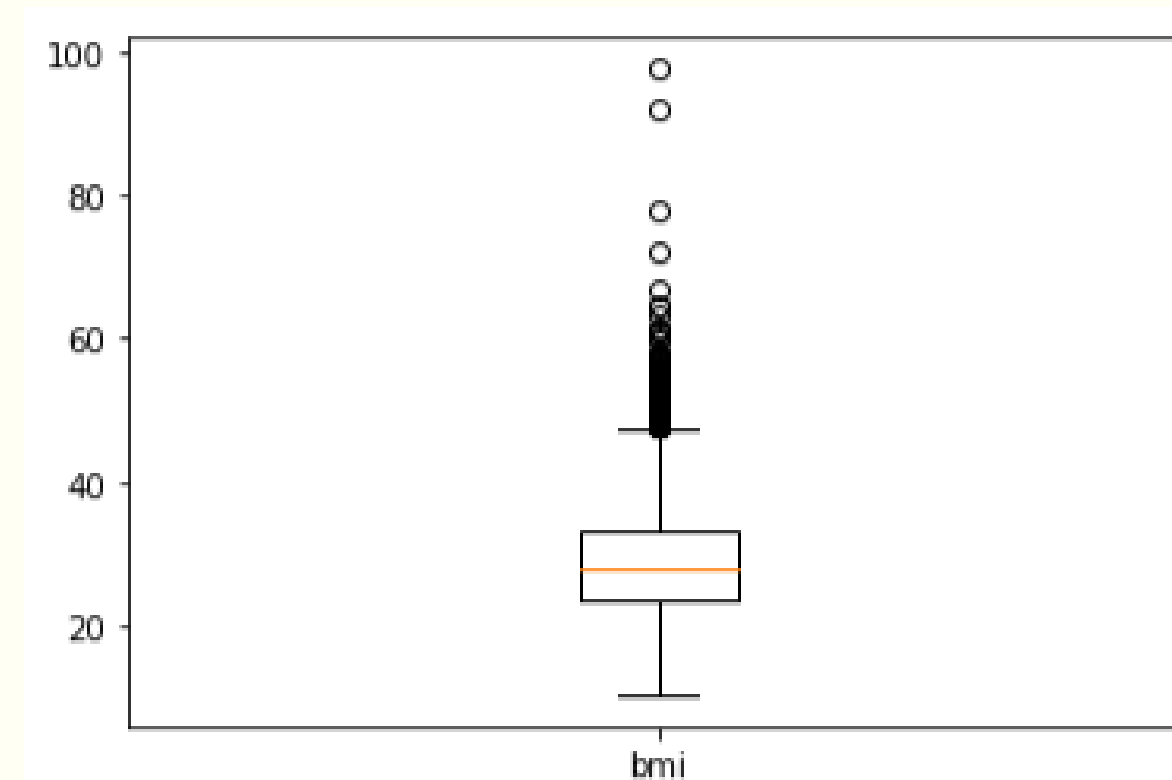
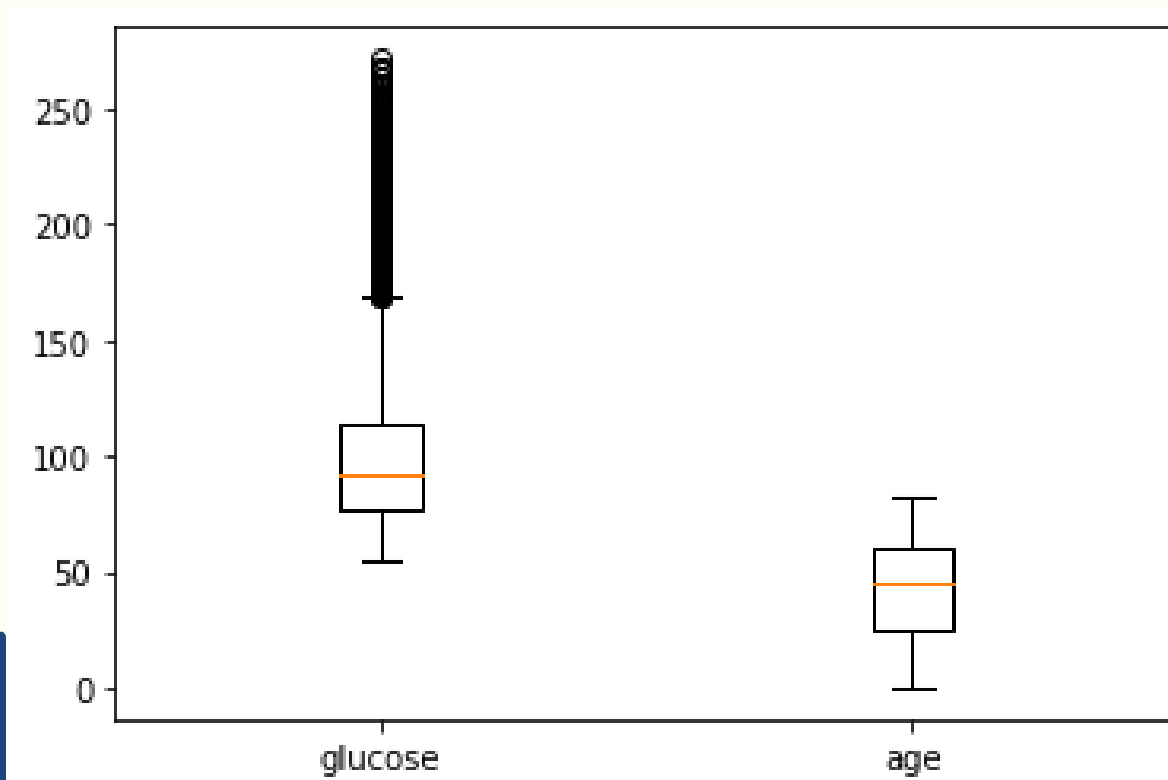
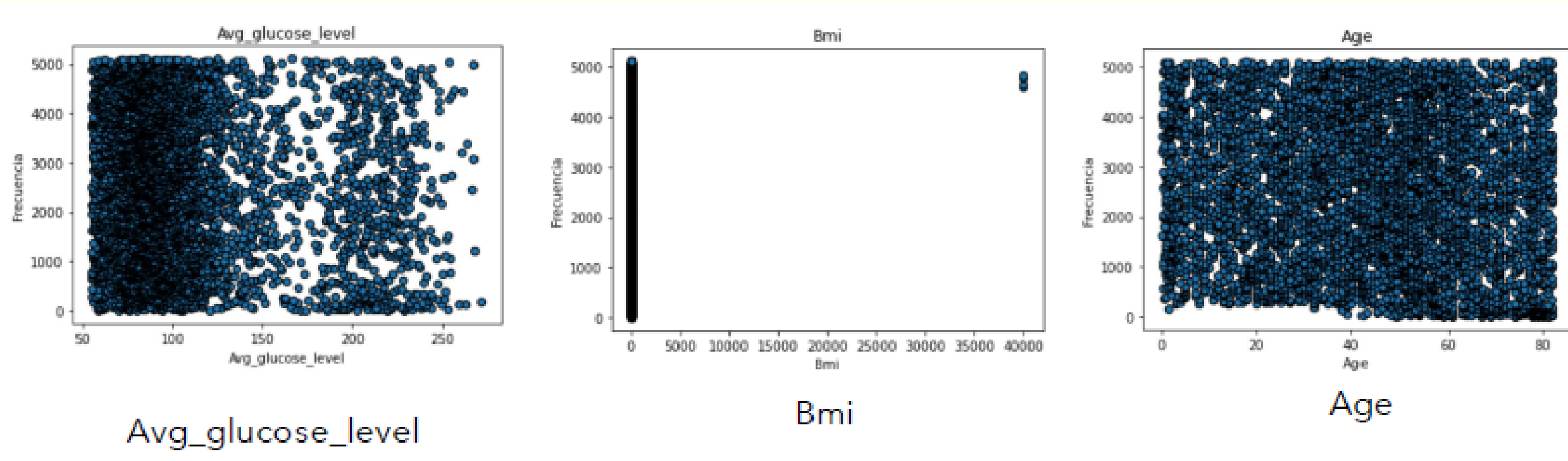


# Correcciones

- Gráfico de cajas en lugar de gráficos de dispersión
- Prueba estadística
- Set de prueba, validación y prueba



# Corrección 1: Gráfico de caja



# Corrección 2: Prueba estadística

**Test de Wilcoxon:** Compara las clasificaciones ordenadas de las muestras. Es adecuado para evaluar la efectividad de técnicas de clasificación, comparando así la precisión promedio de los modelos.

**H0:  $u1 = u2$**  Donde  $u1$  es el puntaje promedio del modelo entrenado con undersampling y  
**H1:  $u1 \neq u2$**   $u2$  es el puntaje promedio del modelo entrenado con oversampling

**Con un nivel de significancia del 5% y un valor p de 0,00195**



**Hay evidencias estadísticas para rechazar la hipótesis nula**



**Conclusion: Se escoge el modelo con underampling**

# Corrección 3: Métricas para los algoritmos

## Set de entrenamiento

## Set de validación

## Set de prueba

**RFC**

Exactitud: 1.0  
Precision: 1.0  
Sensibilidad: 1.0  
F1 Score: 1.0  
AUC: 1.0  
Especificidad: 1.0  
ROC: 1.0

Exactitud: 0.9846  
Precision: 0.9712  
Sensibilidad: 1.0  
F1 Score: 0.9854  
AUC: 0.9839  
Especificidad: 0.9677  
ROC: 0.9838709677419355

Exactitud: 0.9876  
Precision: 0.9764  
Sensibilidad: 1.0  
F1 Score: 0.9881  
AUC: 0.9874  
Especificidad: 0.9747  
ROC: 0.9873551106427818

**SVM**

Exactitud: 0.7691  
Precision: 0.7384  
Sensibilidad: 0.8179  
F1 Score: 0.7761  
AUC: 0.7702  
Especificidad: 0.7224  
ROC: 0.7701704867006225

Exactitud: 0.7894  
Precision: 0.782  
Sensibilidad: 0.8261  
F1 Score: 0.8035  
AUC: 0.7878  
Especificidad: 0.7495  
ROC: 0.7877746610565686

Exactitud: 0.7667  
Precision: 0.7547  
Sensibilidad: 0.8056  
F1 Score: 0.7793  
AUC: 0.7658  
Especificidad: 0.726  
ROC: 0.7658334367973072

Siguientes dos algoritmos

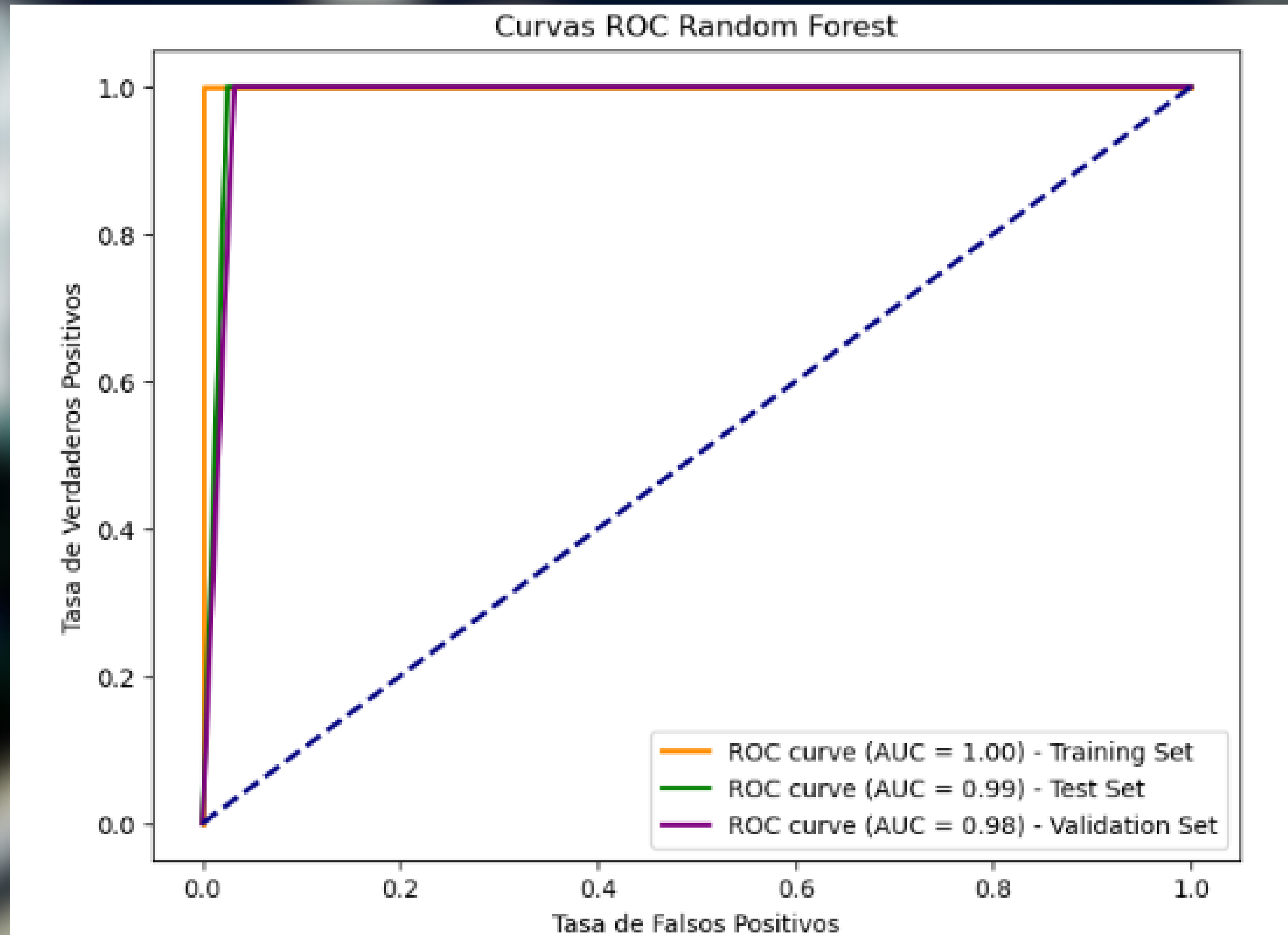
- Regresión Logística: Algoritmo simple que tiene capacidad de manejar muchos datos y es fácil de interpretar.
- Redes Neuronales Artificiales (ANN): Modelo no visto en clase. Capaz de modelar relaciones complejas en los datos y también es capaz de trabajar con grandes cantidades de estos. Este algoritmo usa capas ocultas que permiten aprender características más complejas de los datos.

Continuación 5. Algoritmos y mejor  
técnica de entrenamiento

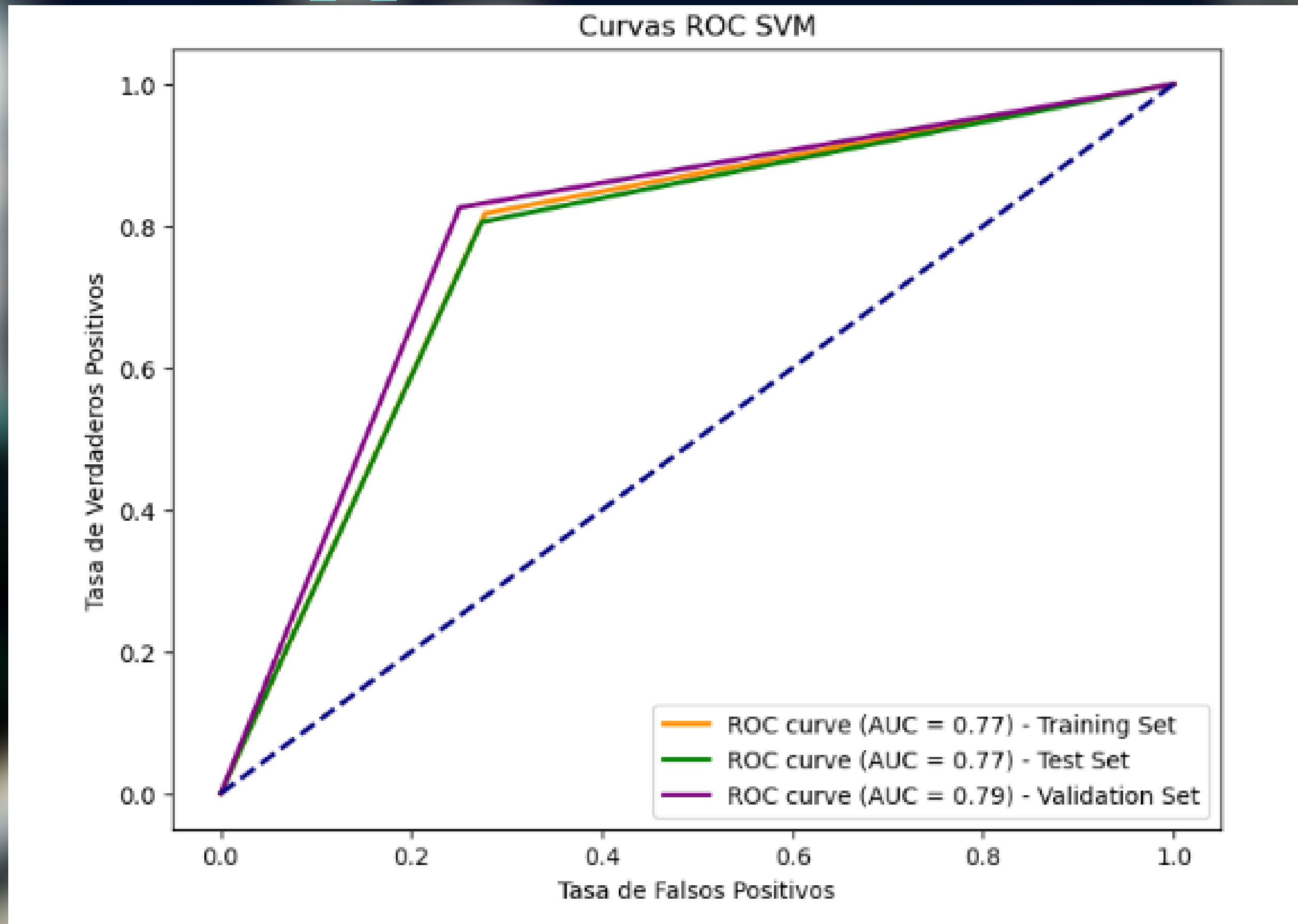


# Métricas de Evaluación y curvas ROC

# ROC Random Forest Classifier



# ROC Support Vector Machines



# Regresión Logística

## Set de entrenamiento

Exactitud: 0.7343  
Precision: 0.7073  
Sensibilidad: 0.7793  
F1 Score: 0.7416  
AUC: 0.7352  
Especificidad: 0.6912  
ROC: 0.7352472410865873

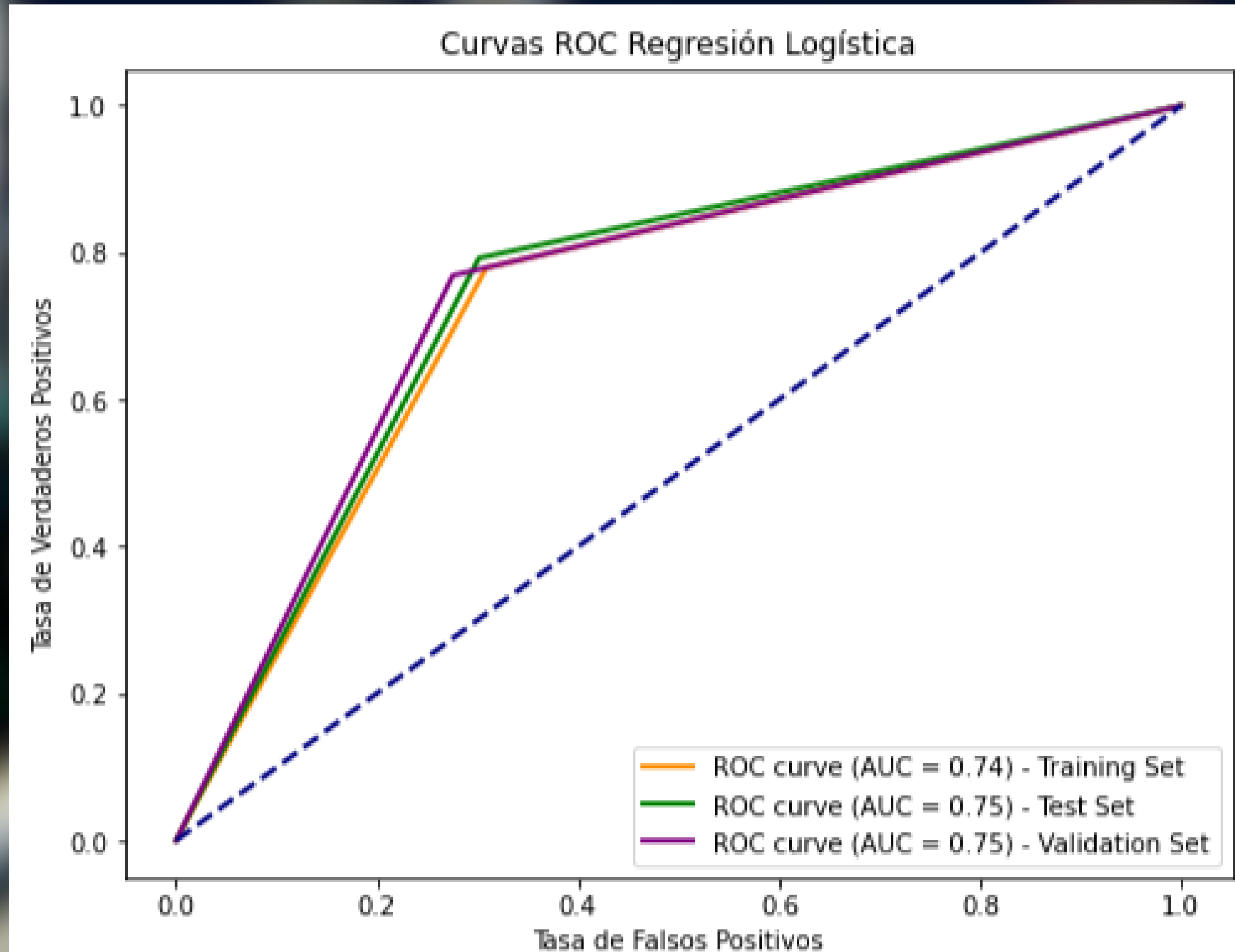
## Set de validación

Exactitud: 0.7477  
Precision: 0.7524  
Sensibilidad: 0.7688  
F1 Score: 0.7605  
AUC: 0.7468  
Especificidad: 0.7247  
ROC: 0.7467529431765056

## Set de prueba

Exactitud: 0.7467  
Precision: 0.7335  
Sensibilidad: 0.7925  
F1 Score: 0.7619  
AUC: 0.7456  
Especificidad: 0.6986  
ROC: 0.7455889859151043

# Regresión Logística: Curvas ROC



# Redes Neuronales Artificiales (ANN)

## Set de entrenamiento

Exactitud: 0.8474  
Precision: 0.8146  
Sensibilidad: 0.8909  
F1 Score: 0.851  
AUC: 0.8483  
Especificidad: 0.8058  
ROC: 0.8483283814374646

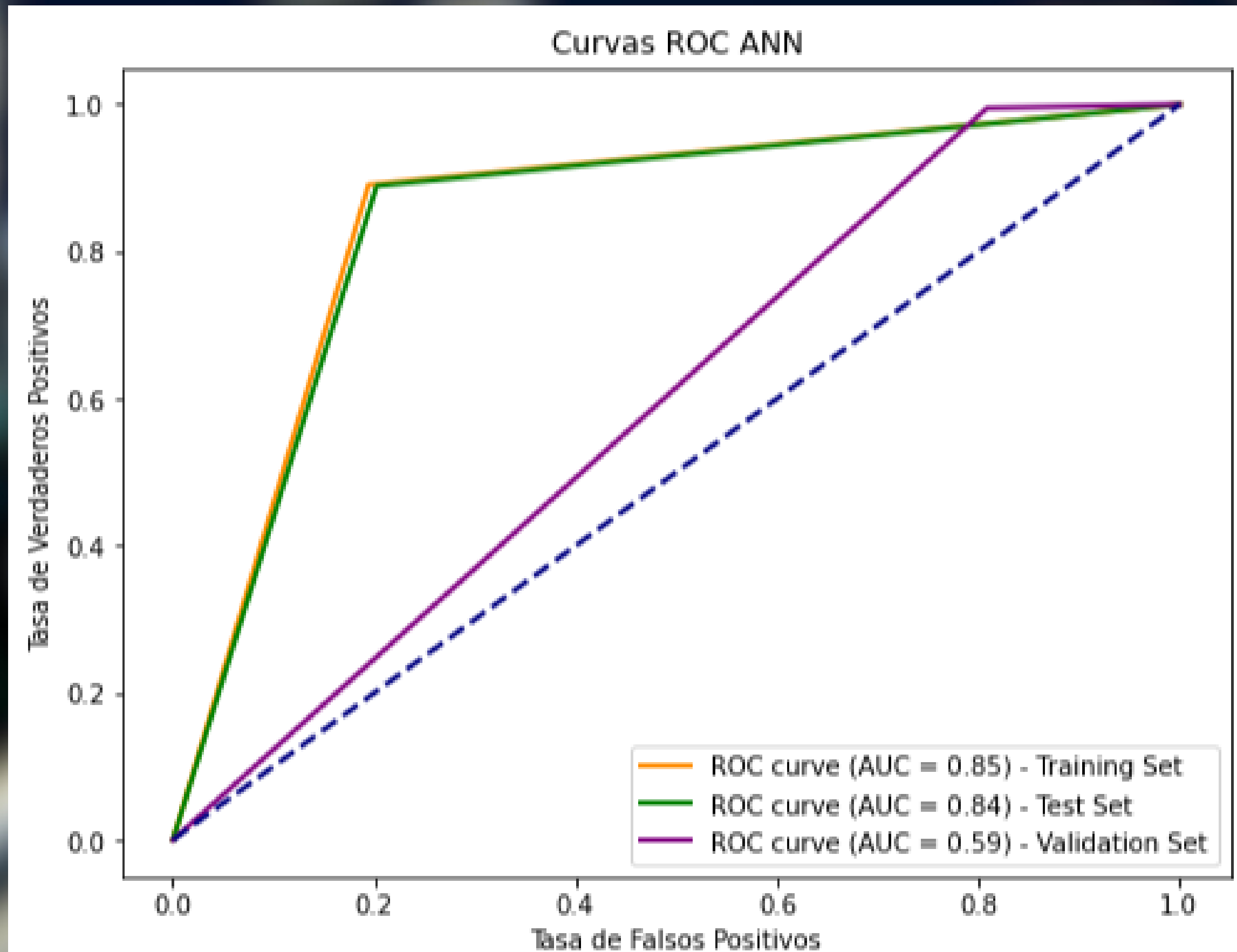
## Set de validación

Exactitud: 0.6102  
Precision: 0.5725  
Sensibilidad: 0.9951  
F1 Score: 0.7268  
AUC: 0.5932  
Especificidad: 0.1914  
ROC: 0.5932285689999575

## Set de prueba

Exactitud: 0.8445  
Precision: 0.8214  
Sensibilidad: 0.8892  
F1 Score: 0.854  
AUC: 0.8435  
Especificidad: 0.7977  
ROC: 0.8434531711442691

# Redes Neuronales Artificiales: Curvas ROC





# Costo de implementación

- Los costos de implementación del modelo están dados por varios factores que varían como: infraestructura, desarrollo e implementación.
- Ventajas:
  - Diagnóstico asistido
  - Detección temprana de afecciones
  - Integración de modelos en el sistema
- Desventajas:
  - Costos elevados de integración
  - Modelo que soluciona una problemática muy específica





# GitHub

- <https://github.com/ggvvmm/Proyecto-Analitica.git>



# Conclusiones

- **Se afirma que el modelo es exitoso, teniendo un porcentaje de predicción del 98% de exactitud y 97% de precisión; esto teniendo en cuenta que se eligió el modelo utilizado con Random Forest Classifier ya que fue el mejor en distintas características**
- **Las mejores variables predictivas que ayudan a la importancia de la variable de respuesta "Stroke" son: la edad (age), el nivel promedio de glucosa (avg\_glucose\_level) e índice de masa corporal (Bmi)**
- **Se prefirió que el modelo sea sensible, ya que, al trabajar con temas de salud, puede ser riesgoso y se prefiere minimizar los falsos negativos**

# Referencias

- Kuo, M. H., Chang, F. L., & Su, M. H. (2018). The benefits and challenges of applying big data in health care: A systematic review. *Journal of nursing research*, 26(3), 174-182.  
[https://journals.lww.com/jnrtwna/Fulltext/2017/12000/The\\_Lived\\_Experience\\_of\\_Gynecologic\\_Cancer.7.aspx](https://journals.lww.com/jnrtwna/Fulltext/2017/12000/The_Lived_Experience_of_Gynecologic_Cancer.7.aspx)
- Johnson, A. E., Stone, D. J., Celi, L. A., & Pollard, T. J. (2018). The MIMIC code repository: enabling reproducibility in critical care research. *Journal of the American Medical Informatics Association*.  
<https://academic.oup.com/jamia/article/25/1/32/4259424>
- Gutierrez-Garcia, J. O. (2021). Datos de Entrenamiento, Validación y Prueba: ¿Cómo crearlos y qué objetivos tienen? Machine Learning. YouTube.  
<https://www.youtube.com/watch?v=vdYzm4xC7mc>

*Gracias!*