



**UNIVERSIDAD POLITÉCNICA SALESIANA  
SEDE GUAYAQUIL  
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**

**DETERMINACIÓN DE LOS FACTORES DE RIESGO DEL CÁNCER DE MAMA  
MEDIANTE APRENDIZAJE AUTOMÁTICO Y EL ÍNDICE SHAP**

Trabajo de titulación previo a la obtención del  
Título de Ingeniero en Ciencias de la Computación

AUTOR: WELLINGTON CRISTÓBAL MIELES SARMIENTO, DANNY ALEXANDER  
BAQUE RODRÍGUEZ

TUTOR: MIGUEL ÁNGEL QUIROZ MARTÍNEZ

Guayaquil – Ecuador

2022

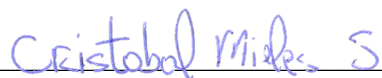
**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE  
TITULACIÓN**

Nosotros, Wellington Cristóbal Mieles Sarmiento con documento de identificación N° 0928109511 y Danny Alexander Baque Rodríguez con documento de identificación N° 0956550230; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

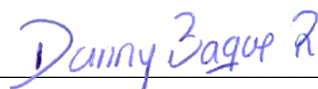
Guayaquil, 3 de marzo del año 2022

Atentamente,

  
\_\_\_\_\_

Wellington Cristóbal Mieles Sarmiento

0928109511

  
\_\_\_\_\_

Danny Alexander Baque Rodríguez

0956550230

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE  
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Nosotros, Wellington Cristóbal Mieles Sarmiento con documento de identificación No. 0928109511 y Danny Alexander Baque Rodríguez con documento de identificación No. 0956550230, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos los autores del Artículo Académico: “DETERMINACIÓN DE LOS FACTORES DE RIESGO DEL CÁNCER DE MAMA MEDIANTE APRENDIZAJE AUTOMÁTICO Y EL ÍNDICE SHAP”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 3 de marzo del año 2022

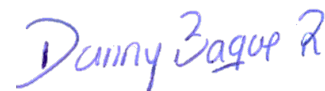
Atentamente,



---

Wellington Cristóbal Mieles Sarmiento

0928109511



---

Danny Alexander Baque Rodríguez

0956550230

## CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Miguel Ángel Quiroz Martínez con documento de identificación N° 0922799655, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DETERMINACIÓN DE LOS FACTORES DE RIESGO DEL CÁNCER DE MAMA MEDIANTE APRENDIZAJE AUTOMÁTICO Y EL ÍNDICE SHAP realizado por Wellington Cristóbal Mieles Sarmiento con documento de identificación N° 0928109511 y por Danny Alexander Baque Rodríguez con documento de identificación N° 0956550230, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 3 de marzo del año 2022

Atentamente,



---

Miguel A. Quiroz Martínez, Msc

0922799655

## **DEDICATORIA**

Dedicamos este trabajo en primer lugar a Dios, por estar en nuestros caminos y permitirnos el haber llegado hasta este momento muy importante en mi vida. A nuestros padres, nuestros pilares fundamentales, siempre confiando y apoyándonos en la vida, a todas las personas de las cuales siempre hemos aprendido y nos motivaron a luchar por nuestras metas.

## **AGRADECIMIENTO**

Agradecemos de manera especial al profesor Miguel Ángel Quiroz Martínez por aceptarnos para realizar el trabajo de titulación bajo su dirección. Su apoyo y confianza en el trabajo. Le agradecemos el habernos guiado siempre con los medios suficientes para llevar a cabo todas las actividades propuestas durante el desarrollo del trabajo de titulación. Muchas gracias, Profesor. Para todos los miembros del Departamento de Ciencias Computacionales: profesores, doctorados y personal administrativo.

## RESUMEN

El cáncer de mama es un tumor maligno que afecta a las personas de todo el mundo, con mayor frecuencia en el sexo femenino, aunque no excluye al sexo masculino. Se encuentra entre los cinco tipos de cáncer más mortíferos, teniendo mayor afluencia en los países menos desarrollados donde el acceso a los programas de salud es más deficiente. Encontrar el mejor algoritmo de machine learning (ML) para una predicción de cáncer de mamá efectiva con la mínima posibilidad de error. Esto permite plantear como objetivo el poder analizar los diferentes algoritmos y seleccionar el que menor error tenga enriqueciendo los algoritmos con el dataset de Breast Cáncer Wisconsin. Se concluyó que el mejor algoritmo es XGBoost con Shap; Utilizamos matrices de confusión para ver el desempeño de los algoritmos, Curva de Roc para conocer el rendimiento global y su precisión dividiendo las predicciones correctas por el total de predicciones.

**Palabras claves:** Factores, Cáncer Mama, Aprendizaje Automático, Índice Shap.

## ABSTRACT

Breast cancer is a malignant tumor that affects people all over the world, more frequently in the female sex, although it does not exclude the male sex. It is among the five deadliest types of cancer, having a greater influx in less developed countries where access to health programs is poorer. Finding the best machine learning (ML) algorithm for effective breast cancer prediction with the least chance of error. This allows us to set the objective of being able to analyze the different algorithms and select the least error that it has, enriching the algorithms with the Breast Cancer Wisconsin data set. It was concluded that the best algorithm is XGBoost with Shap; We use confusion matrices to see the performance of the algorithms, Roc Curve to know the global performance and its accuracy by dividing the correct predictions by the total predictions.

**Key words:** Factors, Breast Cancer, Machine Learning, Index Shap.



## ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN.....	10
2. REVISIÓN DE LITERATURA .....	12
3. METODOLOGÍA.....	13
3.1. Métodos y técnicas de Recopilación de datos empleado .....	13
3.2. Métodos y técnicas de Análisis de datos.....	13
4. RESULTADO .....	15
4.1. Matriz de correlación .....	15
4.2. Matriz de confusión .....	15
4.3. Precisión de los algoritmos .....	16
5. CONCLUSIÓN .....	18
REFERENCIAS.....	19

## 1. INTRODUCCIÓN

El cáncer de mama es el tipo de cáncer más común, con más de 2,2 millones de casos en 2020. El tratamiento del cáncer de mama puede ser muy eficaz, en particular cuando la enfermedad se detecta temprano. (Anderson. B, 2012) En las últimas dos décadas, el número total de personas diagnosticadas con cáncer casi se duplicó, de un estimado de 10 millones en 2000 a 19,3 millones en 2020. Hoy, una de cada 5 personas en todo el mundo desarrollará cáncer durante su vida. Las proyecciones sugieren que la cantidad de personas a las que se les diagnostica cáncer aumentará aún más en los próximos años, y será casi un 50 % más alta en 2040 que en 2020. (Sergey Volkov, 2021) Actualmente la medicina se ha tornado aún más compleja, se dispone de muchas más terapias, medicamentos, exámenes, y procesar tales datos exceden la capacidad de comprensión de la mente humana. (Álvarez Vega et al., 2020) Es necesario encontrar una herramienta que integre todos los datos, reconozca patrones y cree modelo que permita resolver los tiempos de respuesta, disminuir recursos y brindar mejores resultados. (Gui & Chan, 2017)

El aprendizaje automático es una rama de la inteligencia artificial que está presente en muchas áreas, como son: finanzas, mercadotecnia y ciencias sociales para predecir comportamientos desde un conjunto de datos de entrenamiento inicial de características y resultados, un algoritmo aprende cómo las características se relacionan y predicen los resultados. (Char et al., 2018) La capacidad de interpretar correctamente la salida de un modelo de predicción es extremadamente importante. Genera la confianza adecuada del usuario, proporciona información sobre cómo se puede mejorar un modelo y respalda la comprensión del proceso que se está modelando. En algunas aplicaciones, se prefieren modelos simples (por ejemplo, modelos lineales) por su facilidad de interpretación, incluso si pueden ser menos precisos que los complejos. Sin embargo, la creciente disponibilidad de big data ha aumentado los beneficios del uso de modelos complejos, lo que pone de relieve la compensación entre la precisión y la interpretabilidad de la salida de un modelo. (Lundberg et al., n.d.)

La creciente tensión entre la precisión y la interpretabilidad de las predicciones del modelo ha motivado el desarrollo de métodos que ayudan a los usuarios a interpretar las predicciones. El marco SHAP identifica la clase de métodos de importancia de características aditivas (que incluye seis métodos anteriores) y muestra que hay una solución única en esta clase que se adhiere a las propiedades deseables. (Zhao et al., 2020) El hilo de unidad que SHAP teje a través

de la literatura es una señal alentadora de que los principios comunes sobre la interpretación de modelos pueden informar el desarrollo de métodos futuros.

Entre los métodos de detección de cáncer de mama como: Logistic Regression(LR) (Arciniegas et al., 2015), k-Nearest Neighborhood (KNN) (Xabier & Martin, 2015), Decision Tree (DT) (Rocío Erandi Barrientos Martínez et al., 2009), Random Forest (RF) (Fátima Medina Merino & Ismelda Ñique Chacón, 2017), Support Vector Machine (SVM) Classifier(Awad & Khanna, 2015), : Naive Bayes (NB) Classifier (Gutiérrez Esparza et al., 2017), Gradient Boosting (GB)(Zhang & Haghani, 2015), Stochastic Gradient Decent (SGC)(Friedman, 2002), AdaBoost (AB) (Weiming Hu et al., 2008), Extreme Gradient Boosting (XGB)(Chen & He, 2017), CatBoost (CB) (Hancock & Khoshgoftaar, 2020), Light Gradient Boosting( LGB) (Machado et al., 2019), Artificial Neural Networks(ANN)(Basheer & Hajmeer, 2000), SHAP\_XGBoost(Extreme Gradient Boosting – Shap)(Zhao et al., 2020).

Existen técnicas sobre imágenes para análisis de mama como: mamografía, resonancia, ultrasonido y biopsia; la mamografía es la más utiliza en búsqueda de características de cáncer.

La necesidad actual de encontrar el mejor algoritmo de machine learning (ML) para una predicción de cáncer de mamá efectiva con la mínima posibilidad de error. Planteando el objetivo de poder analizar los diferentes algortimos de ML comparándolos para seleccionar el mejor.

## 2. REVISIÓN DE LITERATURA

Items	Algoritmos ML	Ref	Conjunto de datos	Algoritmo seleccionado
1	Metodología propuesta Regresión logística Modelo de KNN	(Arciniegas et al., 2015)	UnbalancedDataset	precisión del 98,24%
2	KNN, LR, DT, RF, SVM, Deep Learning Using Artificial Neural Networks Learning (DL ANNL)	(Xabier & Martin, 2015)	Conjuntos de dataset en una sola tabla	precisión del 95,61%
2	(Gupta & Garg, 2020)	Conjunto de datos sobre el cáncer de mama de Wisconsin	DL ANNL con una precisión del 98,24%	
3	NB, DT, RF, SVM, KNN, Conjunto propuesto	(Sunanda Das, 2019)	Cáncer de mama Wisconsin	Conjunto propuesto con una precisión de 99.28
4	BLR, LDA, QDA, RT, SVM, RF, BA y AB	(Rahman et al., 2020)	Aprendizaje automático UCI	SVM con una precisión del 93,9%
5	ResNet, SVM, ELM y Cross-task extreme ML	(Wang et al., 2020)	Imágenes histopatología del cáncer de Breast	ML extremo de tareas de Cross con una precisión del 96,36%
6	SVM, NB, linear discriminant, quadratic discriminant, LR, KNN	(Singh, 2019)	Edad, índice de masa corporal, glucosa, insulina, leptina, adiponectina y resistina del cáncer de mama	Precisión de KNN de 92.105%
7	Least square, ridge, elastic, RF, gradient boosting, neural net and Douleur Neuropathique	(Juwara et al., 2020)	Cáncer de mama del Hospital General Judío, Montreal, Canadá	Douleur Neuropathique con error cuadrado 1.43

### 3. METODOLOGÍA

#### 3.1. Métodos y técnicas de Recopilación de datos empleado

Se utilizó el dataset Breast Cáncer Wisconsin. Incluye 31 columnas de datos, que se describen a continuación: id, diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, compactness\_mean, concavity\_mean, concave\_points\_mean, symmetry\_mean, fractal\_dimension\_mean, radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, compactness\_se, concavity\_se, concave\_points\_se, symmetry\_se, fractal\_dimension\_se, radius\_worst, texture\_worst, perimeter\_worst, area\_worst, smoothness\_worst, compactness\_worst, concavity\_worst, concave\_points\_worst, symmetry\_worst, fractal\_dimension\_worst.

#### 3.2. Métodos y técnicas de Análisis de datos

##### Coeficiente de correlación de Karl Pearson

Los coeficientes de correlación son medidas que indican la situación relativa de los mismos eventos con respecto a las dos variables, es decir, son la expresión numérica que indica el grado de relación entre las 2 variables y en qué medida están relacionadas. Son números que varían entre los límites +1 y -1. Su magnitud muestra el grado de asociación entre las variables; el valor  $r = 0$  indica que no hay relación entre las variables; los  $\pm 1$  son indicadores de una correcta correlación positiva (cuando X aumenta o disminuye, Y aumenta o disminuye) o negativa (cuando X aumenta o disminuye, Y disminuye o aumenta). (Suárez & Mario, 2011)

##### Matriz de Confusión

Una matriz de confusión, también conocida como matriz de error, es una tabla resumida que se utiliza para evaluar el rendimiento de un modelo de clasificación. El número de predicciones correctas e incorrectas se resumen con los valores de conteo y se desglosan por cada clase. (Terence Shin, 2020)

A continuación, se muestra una imagen de la estructura de una matriz de confusión de 2x2 (ver Figura 1).

		Valores Actuales	
		Si	No
Predicciones	Si	Verdadero Positivo	Falso Positivo
	No	Falso Negativo	Verdadero Negativo

Figura 1. Estructura de matriz de confusión

A continuación, se muestra los pasos que se siguieron para determinar el mejor algoritmo de machine learning para la detección de cáncer de mama:

1. Se importan las librerías (pandas, numpy, seaborn, matplotlib.pyplot).
2. Se importa el dataset (Breast Cáncer Wisconsin) y se hace la visualización de los datos.
3. Eliminación de las columnas de ID y Unnamed.
4. Realización de la matriz de Pearson para saber que datos están correlacionados.
5. Dividimos los datos de entrenamiento y prueba.
6. Utilización de StandarScaler para las variables X\_train y X\_test para estandarizarlas.
7. Utilización de los algoritmos de machine learning para su entrenamiento y sus predicciones.
8. Realización de la matriz de confusión para ver el desempeño de los modelos de machine learning.
9. Realización de la precisión mediante la formula ( $ACURACCY = \frac{PREDICCIONES CORRECTAS}{NUMERO DE PREDICCIONES}$ ).
10. Realización de la curva de ROC para conocer el rendimiento general de los modelos.
11. Realización de una tabla con todos los algoritmos con su precisión.

Plataforma web utilizada para encontrar el mejor modelo

Colaboratory, o "Colab" para abreviar, es un producto de Google Research. Permite a cualquier usuario escribir y ejecutar código arbitrario de Python en el navegador. Es especialmente adecuado para tareas de aprendizaje automático, análisis de datos y educación.

## 4. RESULTADO

### 4.1. Matriz de correlación

La correlación es la fuerza de la relación entre dos atributos; Un coeficiente de correlación está entre -1 y +1; 1 positivo indica una perfecta correlación lineal, y 1 negativo indica una correlación lineal negativa perfecta; cero indica que los atributos no están correlacionados. Para un mayor entendimiento de los valores hay colores que representan rápidamente los datos que están correlacionados (ver Figura2).

- Entre 0 y 0,10: correlación inexistente
- Entre 0,10 y 0,29: correlación débil
- Entre 0,30 y 0,50: correlación moderada
- Entre 0,50 y 1,00: correlación fuerte

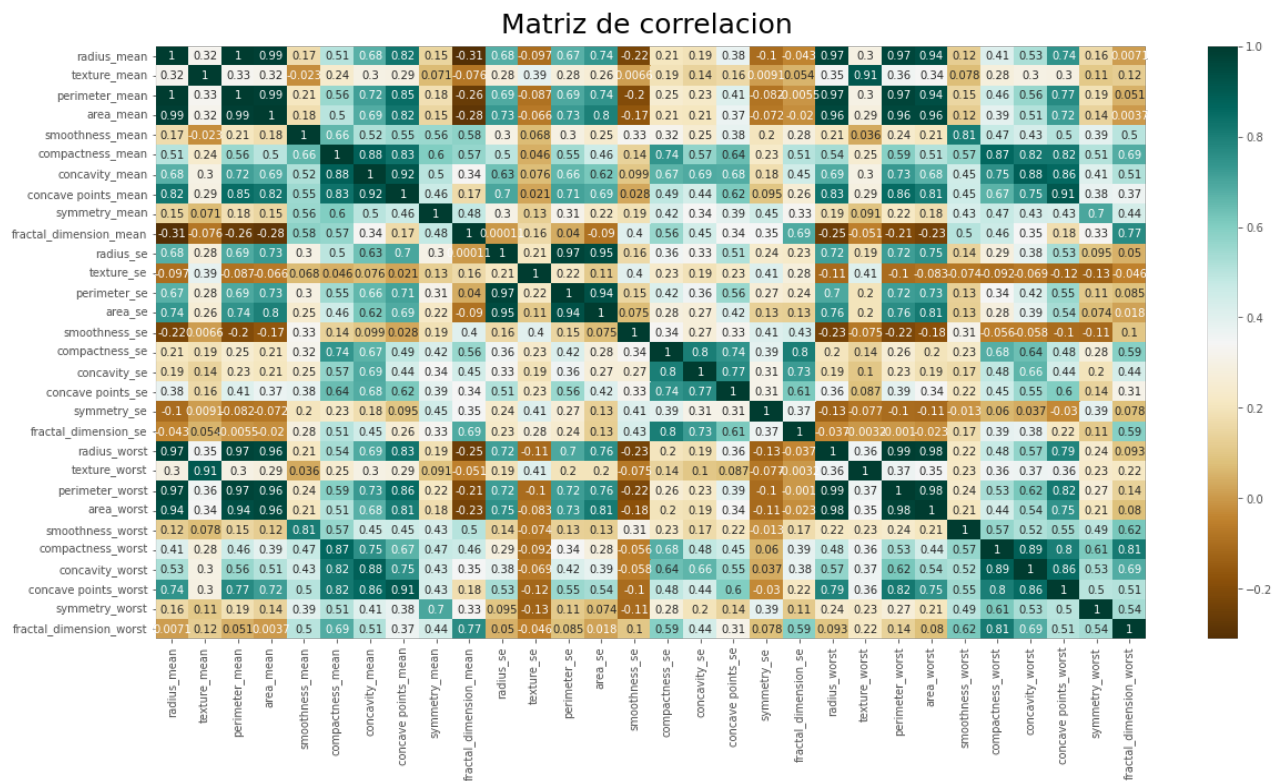


Figura 2. Matriz de correlación

### 4.2. Matriz de confusión

Se obtuvieron los siguientes resultados donde: 1 son tumores malignos y 0 son benignos (ver Figura3).

- 1 en la parte inferior derecha muestra los tumores malignos y arriba de este se muestran los tumores que se clasificaron como malignos, pero en realidad son benignos.
- 0 en la parte superior izquierda muestra los tumores benignos y debajo de este se muestran los tumores que se clasificaron como benignos, pero en realidad son malignos.

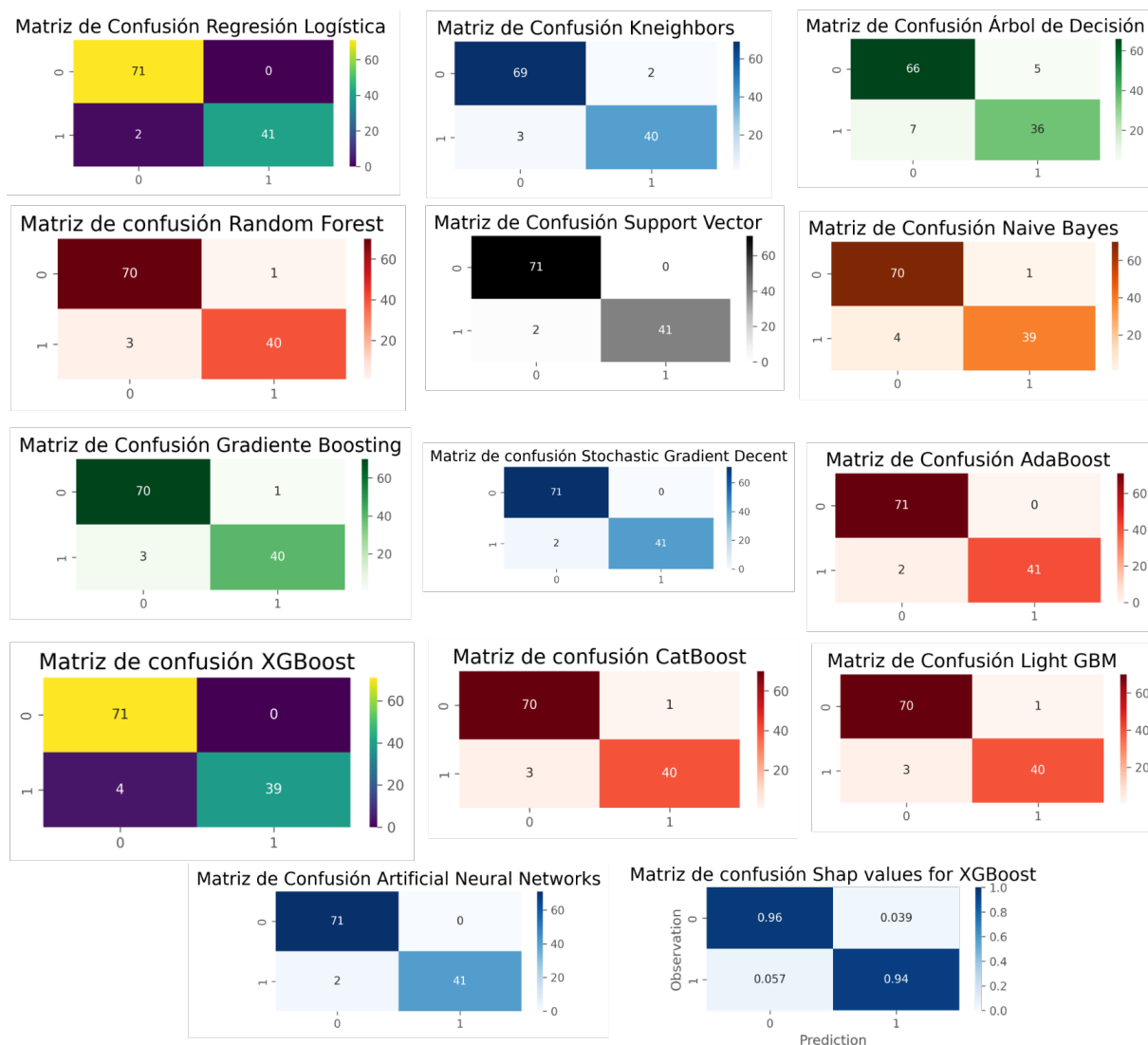


Figura 3 Matrices de confusión

### 4.3. Precisión de los algoritmos

Se obtuvieron las siguientes precisiones de los algoritmos de machine learning, ver Tabla 1.

Ítem	Método	Precisión	Margen de Error
1	Regresión Logística	0,982456140350877	0,017543859649123
2	KNN	0,956140350877193	0,043859649122807
3	Árbol de Decisión	0,912280701754385	0,087719298245615



4	Random Forest	0,964912280701754	0,035087719298246
5	Support Vector Machine	0,982456140350877	0,017543859649123
6	Naive Bayes	0,956140350877193	0,043859649122807
7	Gradient Boosting	0,964912280701754	0,035087719298246
8	Stochastic Gradient Decent	0,956140350877193	0,043859649122807
9	AdaBoost	0,973684210526315	0,026315789473685
10	XGBoost	0,964912280701754	0,035087719298246
11	CatBoost	0,964912280701754	0,035087719298246
12	Light GBM	0,964912280701754	0,035087719298246
13	ANN	0,982456140350877	0,017543859649123
14	SHAP_XGBoost	0,984166617121046	0,015833382878954

*Tabla 1 resultados de cada método*

## 5. CONCLUSIÓN

Se concluyó que todos los algoritmos de machine learning tiene una buena precisión, pero el que menor margen de error tiene es XGBoost con shap dada su matriz de confusión también se puede ver como las predicciones discretas a partir de las probabilidades tienen un gran desempeño con una precisión promedio en torno al 95%.

## REFERENCIAS

- Álvarez Vega, M., Quirós Mora, L. M., & Cortés Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Medica Sinergia*, 5(8), e557. <https://doi.org/10.31434/rms.v5i8.557>
- Anderson. B. (2012, March 26). *Cancer de mama*. <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>.
- Arciniegas, C. P., María Del Pilar, A., & Profesor, V. (2015). *Uso de Regresión Logística para predecir deserción estudiantil temprana*.
- Awad, M., & Khanna, R. (2015). Support Vector Machines for Classification. In *Efficient Learning Machines* (pp. 39–66). Apress. [https://doi.org/10.1007/978-1-4302-5990-9\\_3](https://doi.org/10.1007/978-1-4302-5990-9_3)
- Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1), 3–31. [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3)
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
- Chen, T., & He, T. (2017). *xgboost: eXtreme Gradient Boosting*.
- Fátima Medina Merino, R., & Ismelda Ñique Chacón, C. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Dialnet*, 10, 165–189. <https://www.kaggle.com/primaryobjects/voicegender>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gui, C., & Chan, V. (2017). Machine learning in medicine. *University of Western Ontario Medical Journal*, 86(2), 76–78. <https://doi.org/10.5206/uwomj.v86i2.2060>
- Gupta, P., & Garg, S. (2020). Breast Cancer Prediction using varying Parameters of Machine Learning Models. *Procedia Computer Science*, 171, 593–601. <https://doi.org/10.1016/j.procs.2020.04.064>
- Gutiérrez Esparza, G., Margain Fuentes, L., Canul Reich, J., & Ramírez del Real, T. A. (2017). Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño docente. *RIED. Revista Iberoamericana de Educación a Distancia*, 20(2), 293. <https://doi.org/10.5944/ried.20.2.17717>
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7(1), 94. <https://doi.org/10.1186/s40537-020-00369-8>
- Juwara, L., Arora, N., Gornitsky, M., Saha-Chaudhuri, P., & Velly, A. M. (2020). Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning. *International Journal of Medical Informatics*, 141(October 2019), 104170. <https://doi.org/10.1016/j.ijmedinf.2020.104170>
- Lundberg, S. M., Allen, P. G., & Lee, S.-I. (n.d.). *A Unified Approach to Interpreting Model Predictions*. <https://github.com/slundberg/shap>

- Machado, M. R., Karray, S., & de Sousa, I. T. (2019). LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry. *2019 14th International Conference on Computer Science & Education (ICCSE)*, 1111–1116. <https://doi.org/10.1109/ICCSE.2019.8845529>
- Rahman, M. M., Ghasemi, Y., Suley, E., Zhou, Y., Wang, S., & Rogers, J. (2020). Machine Learning Based Computer Aided Diagnosis of Breast Cancer Utilizing Anthropometric and Clinical Features. *Irbm*. <https://doi.org/10.1016/j.irbm.2020.05.005>
- Rocío Erandi Barrientos Martínez, Nicandro Cruz Ramírez, Héctor Gabriel Acosta Mesa, Ivonne Rabatte Suárez, María del Carmen Gogeoascoechea Trejo, Patricia Pavón León, & Sobeida L. Blázquez Morales. (2009). *Árboles de decisión como herramienta en el diagnóstico médico*. [http://www.soporte.uv.mx/rm/num\\_antteriores/revmedica\\_vol9\\_num2/articulos/arboles.pdf](http://www.soporte.uv.mx/rm/num_antteriores/revmedica_vol9_num2/articulos/arboles.pdf)
- Sergey Volkov. (2021). Breast cancer now most common form of cancer: WHO taking action. <https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action>
- Singh, B. K. (2019). Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: A comparative investigation in machine learning paradigm. *Biocybernetics and Biomedical Engineering*, 39(2), 393–409. <https://doi.org/10.1016/j.bbe.2019.03.001>
- Suárez, I., & Mario, O. (2011). Coeficiente de correlación de Karl Pearson. <http://repositorio.utn.edu.ec/handle/123456789/766>
- Sunanda Das, D. B. (2019). Prediction of Breast Cancer Using Ensemble Learning. *Proceedings of the 2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, 26–28.
- Terence Shin. (2020). Comprensión de la Matriz de Confusión y Cómo Implementarla en Python. *DataSource AI*.
- Wang, P., Song, Q., Li, Y., Lv, S., Wang, J., Li, L., & Zhang, H. (2020). Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. *Biomedical Signal Processing and Control*, 57, 101789. <https://doi.org/10.1016/j.bspc.2019.101789>
- Weiming Hu, Wei Hu, & Maybank, S. (2008). AdaBoost-Based Algorithm for Network Intrusion Detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(2), 577–583. <https://doi.org/10.1109/TSMCB.2007.914695>
- Xabier, M., & Martin, U. (2015). *Aprendizaje de distancias basadas en disimilitudes para el algoritmo de clasificación kNN*. <https://academica-e.unavarra.es/xmlui/bitstream/handle/2454/18430/MemoriaTFGMikelUrizMartin-1.pdf?s>
- Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58, 308–324. <https://doi.org/10.1016/j.trc.2015.02.019>
- Zhao, W., Joshi, T., Nair, V. N., & Sudjianto, A. (2020). *SHAP values for Explaining CNN-based Text Classification Models*.