

Evaluación de algoritmos de clasificación para la predicción de cáncer de mama basados en mamografías e implementación del modelo óptimo en una aplicación web

Alberto Rey Abelaïra

Romina Astrid Rebrij

Carles Ventura Royo

Universitat Oberta de Catalunya

2024

- 1 Introducción
- 2 Objetivos
- 3 Materiales y métodos
- 4 Resultados
- 5 Conclusiones
- 6 Bibliografía

Índice

1 Introducción

2 Objetivos

3 Materiales y métodos

4 Resultados

5 Conclusiones

6 Bibliografía

Contexto y justificación del trabajo

Contexto del trabajo y estado del arte:

Contexto del trabajo y estado del arte:

- 2.3 millones de mujeres diagnosticadas de cáncer de mama en 2020, falleciendo un total de 685 000 personas. (*OMS*)
- Se realizan un 70 % de biopsias innecesarias por el bajo valor predictivo de las mamografías. (*Elter & Schulz-Wendtland (2007)*)
- Importancia de la extracción de características para la determinar la gravedad y el tipo de cáncer. (*Elter & Schulz-Wendtland (2007)*)

Contexto del trabajo y estado del arte:

- 2.3 millones de mujeres diagnosticadas de cáncer de mama en 2020, falleciendo un total de 685 000 personas. (*OMS*)
- Se realizan un 70 % de biopsias innecesarias por el bajo valor predictivo de las mamografías. (*Elter & Schulz-Wendtland (2007)*)
- Importancia de la extracción de características para la determinar la gravedad y el tipo de cáncer. (*Elter & Schulz-Wendtland (2007)*)
- Eficacia de los algoritmos de aprendizaje profundo para interpretar las mamografías. (*McKinney et al. (2020)*)
- Herramientas de garantía de calidad (BIRADS) para evaluar el riesgo de cáncer. (*American College of Radiology*)
- IA ofrece un futuro prometedor en la mejora del diagnóstico y tratamiento para solventar retos como la calidad de datos y la variabilidad interpretativa.

Índice

1 Introducción

2 **Objetivos**

3 Materiales y métodos

4 Resultados

5 Conclusiones

6 Bibliografía

Objetivo principal:

- Desarrollar una aplicación web con el modelo óptimo para la predicción de cáncer de mama a través de características del tumor para tratar de reducir el número de biopsias y pruebas médicas necesarias.

Objetivo principal:

- Desarrollar una aplicación web con el modelo óptimo para la predicción de cáncer de mama a través de características del tumor para tratar de reducir el número de biopsias y pruebas médicas necesarias.

Objetivos específicos:

- Identificar las técnicas de clasificación más adecuadas para la detección de cáncer de mama.
- Determinar el modelo de clasificación más apropiado para el problema, obteniendo una precisión de al menos 80 %.
- Desarrollar una aplicación web interactiva para facilitar y acelerar las fases de diagnóstico y tratamiento.

Índice

- 1 Introducción
- 2 Objetivos
- 3 **Materiales y métodos**
- 4 Resultados
- 5 Conclusiones
- 6 Bibliografía

- Conjunto de datos *Mammographic Mass* (repositorio UCI).

- Conjunto de datos *Mammographic Mass* (repositorio UCI).
- Datos de 961 mamografías entre 2003 y 2006, con 6 variables cada registros:
 - **BIRADS**: Valoración en escala BIRADS de la masa tumoral (0 a 6).
 - **Age**: Edad del paciente (18 a 96 años).
 - **Shape**: Forma de la masa (categórica).
 - **Density**: Densidad de la masa (categórica).
 - **Margin**: Márgenes de la masa (categórica).
 - **Type**: Severidad del tumor (0 para benigno, 1 para maligno).

- Conjunto de datos *Mammographic Mass* (repositorio UCI).
- Datos de 961 mamografías entre 2003 y 2006, con 6 variables cada registros:
 - **BIRADS**: Valoración en escala BIRADS de la masa tumoral (0 a 6).
 - **Age**: Edad del paciente (18 a 96 años).
 - **Shape**: Forma de la masa (categórica).
 - **Density**: Densidad de la masa (categórica).
 - **Margin**: Márgenes de la masa (categórica).
 - **Type**: Severidad del tumor (0 para benigno, 1 para maligno).

	1	2	3	4	5
Shape	Redondeado	Oval	Lobular	Irregular	N/A
Density	Alta	Iso	Baja	Contiene grasa	N/A
Margin	Circunscrito	Microlobulado	Indistinto	Mal definido	Espiculado

Análisis Exploratorio

- Evaluación de la calidad de los datos.
- Análisis descriptivo de las variables.
- Análisis gráfico de las distribuciones y relaciones entre variables.

Análisis Exploratorio

- Evaluación de la calidad de los datos.
- Análisis descriptivo de las variables.
- Análisis gráfico de las distribuciones y relaciones entre variables.

Preprocesado de Datos

- 1 Codificación one-hot para variables categóricas.
- 2 Normalización min-max para la variable *Age*.

Métodos

Preprocesado de datos

Análisis Exploratorio

- Evaluación de la calidad de los datos.
- Análisis descriptivo de las variables.
- Análisis gráfico de las distribuciones y relaciones entre variables.

Preprocesado de Datos

- 1 Codificación one-hot para variables categóricas.
- 2 Normalización min-max para la variable *Age*.

```
> head(trainData)
```

	Age	Shape	Margin	Density	Type
1	67	3	5	3	1
3	58	4	5	3	1
4	28	1	1	3	0
9	57	1	5	3	1
18	59	2	1	3	1
19	54	1	1	3	1

Métodos

Preprocesado de datos

```
> head(trainDataOH)
```

	Age	Shape.1	Shape.2	Shape.3	Shape.4	Margin.1	Margin.2	Margin.3	Margin.4	Margin.5	Density.1	Density.2	Density.3	Density.4	Type
1	0.6233766	0	0	1	0	0	0	0	0	1	0	0	1	0	1
3	0.5064935	0	0	0	1	0	0	0	0	1	0	0	1	0	1
4	0.1168831	1	0	0	0	1	0	0	0	0	0	0	1	0	0
9	0.4935065	1	0	0	0	0	0	0	0	1	0	0	1	0	1
18	0.5194805	0	1	0	0	1	0	0	0	0	0	0	1	0	1
19	0.4545455	1	0	0	0	1	0	0	0	0	0	0	1	0	1

Preprocesado de Datos

- 1 Codificación one-hot para variables categóricas.
- 2 Normalización min-max para la variable *Age*.

```
> head(trainData)
```

	Age	Shape	Margin	Density	Type
1	67	3	5	3	1
3	58	4	5	3	1
4	28	1	1	3	0
9	57	1	5	3	1
18	59	2	1	3	1
19	54	1	1	3	1

Métodos

Entrenamiento de modelos

Algoritmo	Conjunto de Datos	Técnicas y Parámetros
kNN	Ambos	k fijo, k var. y CV
SVM	Procesados	Combinación de kernel y parámetros
Naive Bayes	Ambos	Simple, Laplace y CV
Reg. Log.	Ambos	Simple, Ridge, Lass y CV

Métodos

Entrenamiento de modelos

Algoritmo	Conjunto de Datos	Técnicas y Parámetros
kNN	Ambos	k fijo, k var. y CV
SVM	Procesados	Combinación de kernel y parámetros
Naive Bayes	Ambos	Simple, Laplace y CV
Reg. Log.	Ambos	Simple, Ridge, Lass y CV

- Se evaluó la adaptabilidad de los algoritmos a los conjuntos crudos y procesados.
- Se aplicaron técnicas de regularización y validación cruzada para optimizar los modelos.
- Se utilizó el paquete caret para el entrenamiento y validación, con parámetros específicos para cada algoritmo.
- Se combinan diferentes técnicas para evaluar la robustez de los modelos.

Métodos

Métricas de evaluación del rendimiento

Accuracy (Precisión)

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

Sensibilidad

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

Especificidad

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

Curva ROC - AUC

Relación entre la tasa de verdaderos positivos y tas de falsos positivos

Valor predictivo positivo

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

FDR

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

Índice

1 Introducción

2 Objetivos

3 Materiales y métodos

4 Resultados

5 Conclusiones

6 Bibliografía

Resultados

Análisis descriptivo

Se analizó la calidad de los datos, eliminando registros erróneos o incompletos.

Resultados

Análisis descriptivo

Se analizó la calidad de los datos, eliminando registros erróneos o incompletos.

	0	1	2	3	4	5	6
Benigno	2	0	7	20	365	31	2
Maligno	3	0	0	4	103	285	7

- Se identificó un valor anómalo en BIRADS (55) y se eliminó por inconsistencia.
- Se elimina la incongruencia BIRADS 6 y clase benigna.

Resultados

Análisis descriptivo

Se analizó la calidad de los datos, eliminando registros erróneos o incompletos.

	0	1	2	3	4	5	6
Benigno	2	0	7	20	365	31	2
Maligno	3	0	0	4	103	285	7

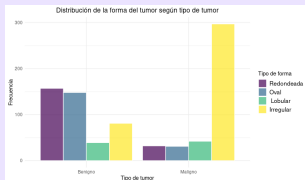
- Se identificó un valor anómalo en BIRADS (55) y se eliminó por inconsistencia.
- Se elimina la incongruencia BIRADS 6 y clase benigna.
- El conjunto final consta de 827 observaciones, equilibradamente distribuidas entre clases benignas y malignas (51.4 % benigna, 48.6 % maligna).

Resultados

Análisis gráfico 1

Resultados

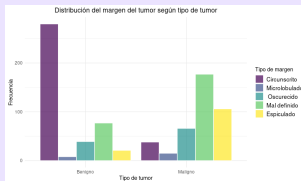
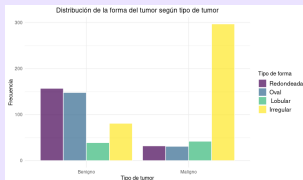
Análisis gráfico 1



- Forma irregular
73 % tumores
malignos.
- Distribución
equitativa en
tumores benignos.

Resultados

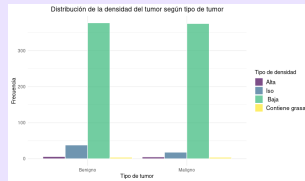
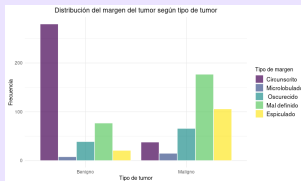
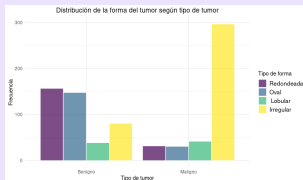
Análisis gráfico 1



- Forma irregular
73 % tumores malignos.
- Distribución equitativa en tumores benignos.
- 65 % margen circunscrito en tumores benignos.
- >50 % tumores malignos espiculado o mal definido.

Resultados

Análisis gráfico 1



- Forma irregular
73 % tumores malignos.

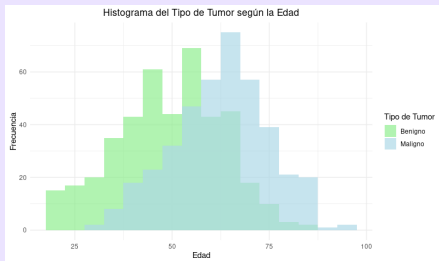
- Distribución equitativa en tumores benignos.

- 65 % margen circunscrito en tumores benignos.
- >50 % tumores malignos espiculado o mal definido.

- Distribución idéntica en ambas clases.
- 90 % densidad baja.

Resultados

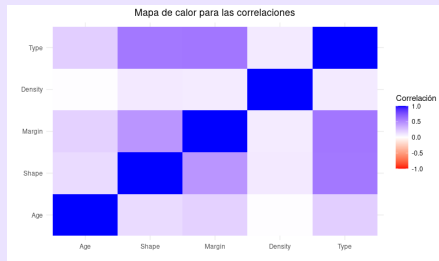
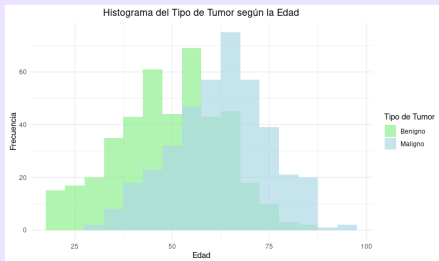
Análisis gráfico 2



- Similar en ambas clases.
- Dependencia del tipo de tumor, siendo benigno en edades más tempranas.

Resultados

Análisis gráfico 2



- Similar en ambas clases.
- Dependencia del tipo de tumor, siendo benigno en edades más tempranas.

- Correlación moderada del tipo con las predictoras.
- Correlación elevada margen - forma.

Resultados

Algoritmo kNN

Modelo	Datos	k Valor	Acc.	Sens.	Espe.
kNN sin CV	Crudos	14	0.743	0.670	0.811
kNN sin CV	One-hot	14	0.825	0.840	0.811
kNN con CV	Crudos	14	0.782	0.750	0.811
kNN con CV	One-hot	25	0.825	0.840	0.811

Resultados

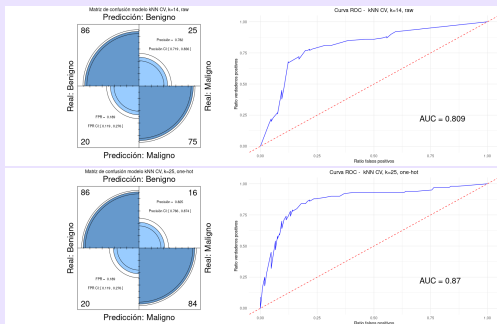
Algoritmo kNN

Modelo	Datos	k Valor	Acc.	Sens.	Espe.
kNN sin CV	Crudos	14	0.743	0.670	0.811
kNN sin CV	One-hot	14	0.825	0.840	0.811
kNN con CV	Crudos	14	0.782	0.750	0.811
kNN con CV	One-hot	25	0.825	0.840	0.811

- Los modelos con one-hot muestran mejor rendimiento, destacando la importancia de los datos adecuados.
- Misma clasificación en ambos modelos one-hot con diferente k . Mejor AUC en modelo con CV.
- Validación cruzada mejora significativamente el rendimiento en datos codificados.
- Mejor rendimiento $k = 25$ con validación cruzada y codificación one-hot.

Resultados

kNN - Gráficas rendimiento



- Importancia de la selección de k , con $k = 25$ ofreciendo la mejor adaptabilidad.
- Mejor rendimiento con datos procesados indica una mejor interpretación numérica por parte del algoritmo.
- La validación cruzada mejora la generalización del modelo, reduciendo el sobreajuste.

Resultados

Algoritmo SVM

Algoritmo	Kernel	CV	C	σ	Acc.	Sens.	Espe.
SVM	Lineal	No	1	NA	0.830	0.860	0.802
SVM	Lineal	Sí	0.1	NA	0.816	0.820	0.811
SVM	Radial	No	0.25	0.22	0.796	0.760	0.830
SVM	Radial	Sí	0.2	0.2	0.786	0.720	0.849

Resultados

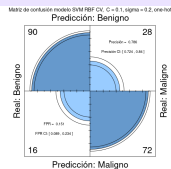
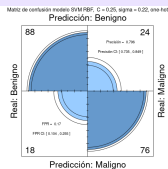
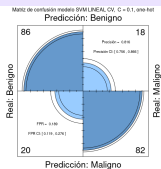
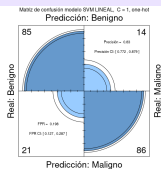
Algoritmo SVM

Algoritmo	Kernel	CV	C	σ	Acc.	Sens.	Espe.
SVM	Lineal	No	1	NA	0.830	0.860	0.802
SVM	Lineal	Sí	0.1	NA	0.816	0.820	0.811
SVM	Radial	No	0.25	0.22	0.796	0.760	0.830
SVM	Radial	Sí	0.2	0.2	0.786	0.720	0.849

- Diferencia rendimiento según el tipo de kernel.
- Poco equilibrio entre sensibilidad y especificidad.
- Los modelos más complejos no interpretan mejor los datos.

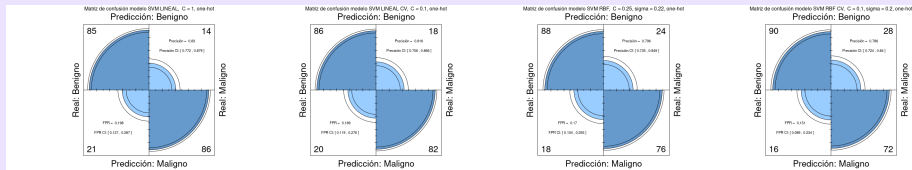
Resultados

SVM - Matrices de confusión



Resultados

SVM - Matrices de confusión



- Las matrices de confusión revelan una variación notable en el equilibrio entre sensibilidad y especificidad.
- Los modelos con kernel radial tienden a aumentar la cantidad de falsos negativos y reducir los verdaderos negativos.
- Los modelos con kernel lineal aumentan la cantidad de falsos positivos.
- Se manifiesta un problema para la predicción clínica.

Resultados

Algoritmo Naive Bayes

Modelo	Datos	Lapl.	Kernel	Acc.	Sens.	Espe.
NB sin CV	Crudos	0	NA	0.782	0.770	0.792
NB con CV	One-hot	0	TRUE(2)	0.811	0.770	0.849
NB sin CV	Crudos	0	NA	0.757	0.800	0.717
NB con CV	One-hot	0	TRUE(1)	0.806	0.850	0.764

Resultados

Algoritmo Naive Bayes

Modelo	Datos	Lapl.	Kernel	Acc.	Sens.	Espe.
NB sin CV	Crudos	0	NA	0.782	0.770	0.792
NB con CV	One-hot	0	TRUE(2)	0.811	0.770	0.849
NB sin CV	Crudos	0	NA	0.757	0.800	0.717
NB con CV	One-hot	0	TRUE(1)	0.806	0.850	0.764

- Variabilidad significativa en el rendimiento dependiendo de la configuración.
- Mejora notable con el uso de validación cruzada y funciones kernel.
- Necesidad de mejora en el equilibrio sensibilidad - especificidad, a pesar de buenas curvas ROC.

Resultados

Algoritmo Regresión Logística

Modelo	Datos	Características	Acc.	Sens.	Espe.
RL simple	Crudos	NA	0.811	0.800	0.821
RL sin CV	One-hot	Mat. reducida	0.816	0.810	0.821
RL CV	One-hot	Mat. reducida	0.835	0.840	0.830

Resultados

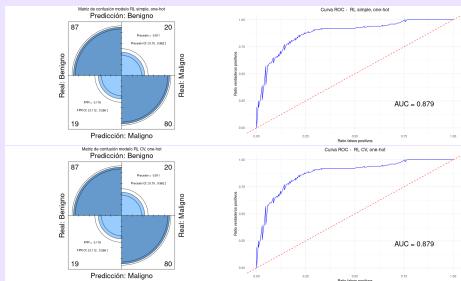
Algoritmo Regresión Logística

Modelo	Datos	Características	Acc.	Sens.	Espe.
RL simple	Crudos	NA	0.811	0.800	0.821
RL sin CV	One-hot	Mat. reducida	0.816	0.810	0.821
RL CV	One-hot	Mat. reducida	0.835	0.840	0.830

- Optimización de parámetros y validación cruzada mejoran el rendimiento en todos los modelos.
- Buen equilibrio entre complejidad del modelo y precisión predictiva.

Resultados

Gráficas de rendimiento - Regresión Logística



- **Modelo sin CV:** Buen rendimiento, con margen de mejora para identificar casos negativos.
- **Modelo con CV:** La CV mejora el balance entre clases, con una mayor precisión.
- **Comparación:** Rendimiento más robusto y estable del modelo con CV. Curvas ROC similares pero mejor número de VP y FN.

- **Punto clave:** El enfoque principal es reducir falsos positivos y negativos, con balance equitativo entre sensibilidad y especificidad.

- **Punto clave:** El enfoque principal es reducir falsos positivos y negativos, con balance equitativo entre sensibilidad y especificidad.
- **Variabilidad Naive Bayes y SVM:** Muestran una variabilidad significativa en su rendimiento, con una tendencia a favorecer una clase sobre la otra, lo que los hace menos ideales para este contexto clínico específico.

- **Punto clave:** El enfoque principal es reducir falsos positivos y negativos, con balance equitativo entre sensibilidad y especificidad.
- **Variabilidad Naive Bayes y SVM:** Muestran una variabilidad significativa en su rendimiento, con una tendencia a favorecer una clase sobre la otra, lo que los hace menos ideales para este contexto clínico específico.
- **Superioridad de la Regresión Logística:** Son más consistentes en la relación deseada, lo que los hace más viables para el problema.

- **Punto clave:** El enfoque principal es reducir falsos positivos y negativos, con balance equitativo entre sensibilidad y especificidad.
- **Variabilidad Naive Bayes y SVM:** Muestran una variabilidad significativa en su rendimiento, con una tendencia a favorecer una clase sobre la otra, lo que los hace menos ideales para este contexto clínico específico.
- **Superioridad de la Regresión Logística:** Son más consistentes en la relación deseada, lo que los hace más viables para el problema.
- **Importancia CV y preprocesado:** La validación cruzada y una adecuada codificación de los datos (como one-hot) emergen como factores clave para mejorar el rendimiento del modelo.

Modelo óptimo y conclusión

- Regresión Logística sobresale en balancear sensibilidad y especificidad.
- El modelo con datos procesados y validación cruzada emerge como el más adecuado.

Modelo óptimo y conclusión

- Regresión Logística sobresale en balancear sensibilidad y especificidad.
- El modelo con datos procesados y validación cruzada emerge como el más adecuado.
- Alta sensibilidad y especificidad, crucial en el contexto clínico del cáncer de mama.
- Importancia de un enfoque meticuloso en la selección y preprocesamiento de datos.

Modelo óptimo y conclusión

- Regresión Logística sobresale en balancear sensibilidad y especificidad.
- El modelo con datos procesados y validación cruzada emerge como el más adecuado.
- Alta sensibilidad y especificidad, crucial en el contexto clínico del cáncer de mama.
- Importancia de un enfoque meticuloso en la selección y preprocesamiento de datos.

Modelo Recomendado para Cada Algoritmo:

Algoritmo	Datos/CV	Parámetros Óptimos
kNN	One-hot, CV	$k = 25$
SVM	One-hot, CV	Lineal/Radial
RL	One-hot, CV	Validación Cruzada
NB	One-hot, CV	Laplace/Kernel

Modelo óptimo y conclusión

- Regresión Logística sobresale en balancear sensibilidad y especificidad.
- El modelo con datos procesados y validación cruzada emerge como el más adecuado.
- Alta sensibilidad y especificidad, crucial en el contexto clínico del cáncer de mama.
- Importancia de un enfoque meticuloso en la selección y preprocesamiento de datos.

Modelo Recomendado para Cada Algoritmo:

Algoritmo	Datos/CV	Parámetros Óptimos
kNN	One-hot, CV	$k = 25$
SVM	One-hot, CV	Lineal/Radial
RL	One-hot, CV	Validación Cruzada
NB	One-hot, CV	Laplace/Kernel

Conclusión: La Regresión Logística con datos procesados y validación cruzada es ideal en un contexto para la detección precisa y minimización de intervenciones innecesarias en el cáncer de mama.

Resultados

AppWeb. Pestaña información

Aplicación para la predicción de cáncer de mama a través de características de mamografías

Información

Clasificación

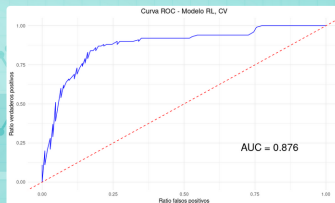
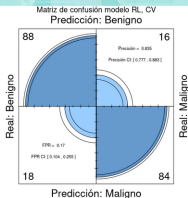
Resultados

Aquí encontrarás información importante sobre la aplicación y el modelo implementado

Esta aplicación es el resultado de un proyecto de investigación desarrollado como parte del Trabajo de Fin de Máster en Bionformática y Bioestadística en la UOC en el área de Estadística y aprendizaje automático. Su objetivo es demostrar la aplicación práctica de técnicas avanzadas de análisis de datos en el campo de la predicción de cáncer de mama.

[Haz clic aquí para acceder al repositorio del proyecto y obtener más detalles sobre el TFM.](#)

El modelo de clasificación implementado en esta aplicación utiliza la clasificación basada en Regresión Logística para predecir si una tumoración mamaria es benigna o maligna. Este modelo fue entrenado y validado utilizando un conjunto de datos del repositorio [UCI](#), recopilados en el estudio *The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process* (doi:10.1118/1.2786864). La predicción que se obtiene con este modelo tiene una precisión del 83.5%, una sensibilidad y especificidad del 84% y 83% respectivamente, y un valor predictivo positivo del 80.4%. En las imágenes siguientes se muestran los gráficos de rendimiento del modelo:



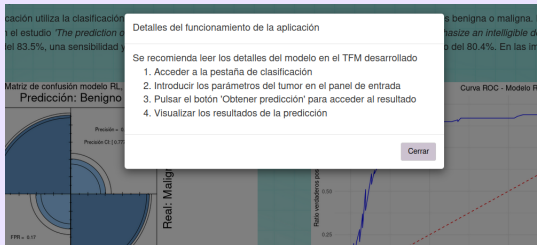
Resultados

AppWeb. Pestaña información 2

Haz clic [aquí](#) para detalles sobre cómo obtener una predicción. En la siguiente tabla puedes consultar los parámetros de entrada.

Variable	Descripción	Tipo	Rango
Edad	Edad del sujeto	Númerica	18 - 96
Forma	Forma de la tumoración	Categórica	1 - 4
Margen	Margen de la tumoración	Categórica	1 - 5
Densidad	Densidad de la tumoración	Categórica	1 - 4

© 2024, Alberto Rey Abelaíra - albertorey@uoc.edu



Resultados

AppWeb. Pestaña clasificación y resultados

Panel de entrada de datos 1

Resultados

AppWeb. Pestaña clasificación y resultados

Clasificación

Edad

Forma del tumor (BIRADS)

Selecciona una opción

Margen del tumor (BIRADS)

Selecciona una opción

Densidad del tumor (BIRADS)

Selecciona una opción

Obtener clasificación

Panel de entrada de datos 1

Clasificación

Edad

19

Forma del tumor (BIRADS)

Oval (2)

Margen del tumor (BIRADS)

Mal definido (4)

Densidad del tumor (BIRADS)

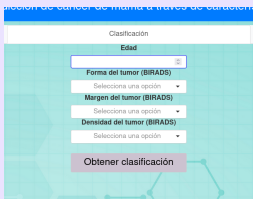
Iso (2)

Obtener clasificación

Panel de entrada de datos 2

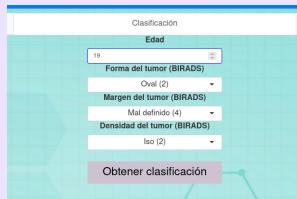
Resultados

AppWeb. Pestaña clasificación y resultados



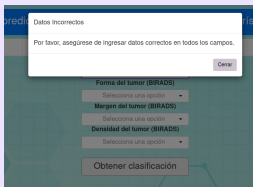
Panel de entrada de datos 1. Interfaz de usuario para la clasificación de tumores. Incluye campos para: Edad (campo de texto), Forma del tumor (BIRADS) (lista desplegable), Margen del tumor (BIRADS) (lista desplegable), Densidad del tumor (BIRADS) (lista desplegable), y un botón 'Obtener clasificación'.

Panel de entrada de datos 1



Panel de entrada de datos 2. Interfaz de usuario para la clasificación de tumores. Incluye campos para: Edad (campo de texto con valor 19), Forma del tumor (BIRADS) (lista desplegable con valor Oval (2)), Margen del tumor (BIRADS) (lista desplegable con valor Mal definido (4)), Densidad del tumor (BIRADS) (lista desplegable con valor Iso (2)), y un botón 'Obtener clasificación'.

Panel de entrada de datos 2

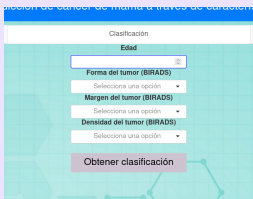


Datos erróneos. Pantalla de error que indica: 'Datos incorrectos. Por favor, asegúrese de ingresar datos correctos en todos los campos.' con un botón 'Cerrar'. El fondo muestra la interfaz de usuario para la clasificación de tumores.

Datos erróneos

Resultados

AppWeb. Pestaña clasificación y resultados



Aplicación para la predicción de cáncer de mama a través de resultados de clasificación

Clasificación

Edad

Forma del tumor (BIRADS)

Selecciona una opción

Margen del tumor (BIRADS)

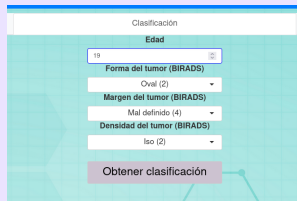
Selecciona una opción

Densidad del tumor (BIRADS)

Selecciona una opción

Obtener clasificación

Panel de entrada de datos 1



Clasificación

Edad

19

Forma del tumor (BIRADS)

Oval (2)

Margen del tumor (BIRADS)

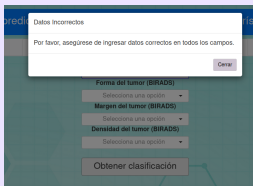
Mal definido (4)

Densidad del tumor (BIRADS)

Iso (2)

Obtener clasificación

Panel de entrada de datos 2



Datos incorrectos

Por favor, asegúrese de ingresar datos correctos en todos los campos.

Cerrar

Forma del tumor (BIRADS)

Selecciona una opción

Margen del tumor (BIRADS)

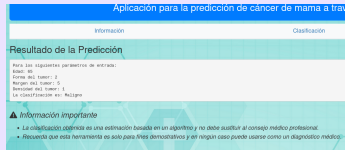
Selecciona una opción

Densidad del tumor (BIRADS)

Selecciona una opción

Obtener clasificación

Datos erróneos



Aplicación para la predicción de cáncer de mama a través de resultados de clasificación

Información

Clasificación

Resultado de la Predicción

Para los siguientes parámetros de entrada:

Edad: 19

Forma del tumor: 2

Margen del tumor: 5

Densidad del tumor: 1

La clasificación es: **Benigno**

Información importante

- La clasificación obtenida es una estimación basada en un algoritmo y no debe sustituir al consejo médico profesional.
- Recuerda que esta herramienta es solo para fines demostrativos y en ningún caso puede usarse como un diagnóstico médico.

Resultado de la predicción

Índice

1 Introducción

2 Objetivos

3 Materiales y métodos

4 Resultados

5 Conclusiones

6 Bibliografía

Conclusiones

Conclusiones generales

- ✓ **Papel clave del tratamiento de los datos:** La codificación one-hot y la normalización de la edad son esenciales para mejorar el rendimiento de los modelos predictivos.
- ✓ **Selección óptima del modelo:** La regresión logística se destaca por su precisión y equilibrio en la detección de tumores en el contexto clínico.
- ✓ **Importancia del uso de la tecnología en medicina:** Visible a través de la aplicación web interactiva.
- ✓ **Eficacia en la predicción de cáncer de mama:** Los resultados obtenidos reflejan la aplicabilidad y eficacia de los métodos de aprendizaje automático en la detección de este cáncer.

Conclusiones

Trabajos futuros

- **Alternativas en el conjunto de datos**
- **Evaluación clínica y validación**
- **Incorporación de nuevas técnicas de análisis y variables.**

Índice

1 Introducción

2 Objetivos

3 Materiales y métodos

4 Resultados

5 Conclusiones

6 Bibliografía

Bibliografía (1)

- [1] Organización Mundial de la Salud. (s.f.). *Cáncer de mama*. World Health Organization. <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>
- [2] Budh, D. P., & Sapra, A. (2022). *Breast Cancer Screening*. [Updated 2022 Oct 22]. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls.
- [3] Elter, M. R., & Schulz-Wendtland, T. W. (2007). *The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process*. *Med. Phys.*, **34**, 4164–4172.
- [4] American College of Radiology (ACR) (2003). *Breast Imaging Reporting and Data System Atlas (BIRADS Atlas)*. Reston, Va: American College of Radiology.
- [5] Naciones Unidas. (s/f). *Objetivos de Desarrollo Sostenible (ODS) - Naciones Unidas*. Recuperado de <https://www.un.org/sustainabledevelopment/es/sustainable-development-goals/>

Bibliografía (2)

- [1] Lantz, B. (2015). *Machine Learning with R*. Birmingham: PACK Publishing.
- [2] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. USA: Springer.
- [3] Harrell, F. E., Jr. (2001). *Regression Modeling Strategies*. USA: Springer.
- [4] Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. USA: Springer.
- [5] Hastie, T., et al. (2013). *An Introduction to Statistical Learning: with applications in R*. USA: Springer.
- [6] Zheng, Alice y Casari, Amanda. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media. ISBN: 978-1491953242.
- [7] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Bibliografía (3)

- [1] Simonyan, K. y Zisserman, A. (2014).
Very Deep Convolutional Networks for Large-Scale Image Recognition.
arXiv preprint arXiv:1409.1556.
- [2] McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., et al. (2020).
International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89–94.
- [3] Bennett, K., & Campbell, C. (2000, diciembre). *Support Vector Machines: Hype or Hallelujah?* *ACM SIGKDD Explorations Newsletter, 2*, 1-13.
<https://doi.org/10.1145/380995.380999> R package version 1.7.4.9002.

Bibliografía (4)

- [1] Elter, M. (2007). *Mammographic Mass*. UCI Machine Learning Repository. <https://doi.org/10.24432/C53K6Z>
- [2] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. <https://www.R-project.org/>
- [3] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020. <http://www.rstudio.com/>.
- [4] Max Kuhn.
Building Predictive Models in R Using the caret Package.
Journal of Statistical Software, 28(5):1–26, 2008.
DOI: 10.18637/jss.v028.i05.
<https://www.jstatsoft.org/index.php/jss/article/view/v028i05>.
- [5] Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., ... Borges, B. (2023). *Shiny: Web Application Framework for R*. R package version 1.7.4.9002.

¡Gracias!

Alberto Rey Abelaira