

COMPARACIÓN DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO SUPERVISADO APLICADAS A DATOS CARDIOLÓGICOS

Nicolás Seivane¹, Andrea Alejandra Rey^{1,2}

¹Laboratorio de Investigación y Desarrollo Experimental en Computación

Universidad Nacional de Hurlingham

²CPSI, Universidad tecnológica Nacional, Regional Buenos Aires

seivanenicolas@gmail.com, andrea.rey@unahur.edu.ar

RESUMEN

En este trabajo se realiza un análisis comparativo de técnicas de Aprendizaje Automático Supervisado aplicadas a datos cardiológicos para evaluar la clasificación de riesgo en pacientes de poseer una afección cardíaca. Se implementaron y evaluaron los algoritmos Naïve Bayes (NB) [1], Regresión Logística (RF) [2], Random Forest (RF) [3] y Máquinas de Vectores de Soporte (SVM) [4]. El estudio abarca tanto un caso binario de insuficiencia cardíaca como uno multiclase de cardiotocografía fetal. Los resultados indican que el método con mejor desempeño general es Random Forest, logrando una alta capacidad de generalización en ambos escenarios.

Palabras clave: Aprendizaje Automático, Cardiología, Clasificación, Aprendizaje Supervisado.

CONTEXTO

Esta investigación se inserta en la línea de ??? del Laboratorio de Investigación y Desarrollo Experimental en Computación (LIDEC) de la Universidad Nacional de Hurlingham (UNAHUR). El proyecto busca desarrollar...

INTRODUCCIÓN

Las enfermedades cardiovasculares representan la principal causa de muerte a nivel mundial, cobrando aproximadamente 17,9 millones de vidas al año, lo que equivale al 31 % de las defunciones globales. El proceso de diagnóstico tradicional, aunque efectivo, puede ser extenso y depende del análisis de parámetros clínicos estructurados, señales cardíacas (como el ECG) e imágenes médicas. Herramientas como el electrocardiograma, las pruebas de esfuerzo y los ecocardiogramas son el estándar clínico actual, pero requieren una interpretación obligatoria experta que puede beneficiarse de la agilidad de los modelos computacionales de aprendizaje automático.

El objetivo central de este estudio es identificar una técnica de aprendizaje automático óptima que permita realizar predicciones precisas sobre la probabilidad de insuficiencia cardíaca. Se exploran no solo casos de clasificación binaria (enfermo/sano), sino también multiclase mediante la cardiotocografía (CTG), que evalúa el bienestar fetal. Basándose en el estado del arte [5] [6], que destaca el uso de modelos como SVM y Naïve Bayes para la detección de arritmias y cardiopatías, este trabajo busca validar estas técnicas en un entorno comparativo con *datasets* reales bajo métricas de calidad estándar.

Diversos estudios han demostrado la efectividad moderada y bajo escrutinio médico de algorit-

mos como RF, SVM y los métodos de ensamble en problemas de clasificación médica. En particular, RF ha mostrado un alto desempeño debido a su capacidad para modelar relaciones no lineales y reducir el sobreajuste mediante la combinación de múltiples árboles de decisión. Asimismo, técnicas de interpretabilidad como la importancia por permutación permiten identificar las variables más influyentes, contribuyendo a la comprensión del modelo, siendo estas las más importantes del modelo, no del estudio, eso lo debe determinar un experto.

En este trabajo se utilizaron los conjuntos de datos *HeartFailure* [7], compuesto por 918 registros correspondientes a un problema de clasificación binaria, y *Cardiotocography* [8], con 2115 registros correspondientes a un problema de clasificación multiclase. Ambos contienen variables clínicas relevantes utilizadas para el análisis y predicción de condiciones cardiovasculares.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El eje central de la investigación es la evaluación del rendimiento de clasificadores supervisados bajo diferentes configuraciones de hiperparámetros.

- **Objetivo General:** Comparar el desempeño de diversas técnicas de Aprendizaje Automático Supervisado sobre conjuntos de datos cardiológicos.
- **Objetivos Específicos:**
 - 1. Evaluar modelos en escenarios de clasificación binaria y multiclase.
 - 2. Identificar las características clínicas más influyentes mediante importancia por permutación.
 - 3. Determinar el equilibrio óptimo entre tiempo de cómputo y precisión.

METODOLOGÍA

A partir de los conjuntos de datos mencionados, se realizó un proceso de preparación que incluyó análisis exploratorio, verificación de valores faltantes, consistencia de tipos de datos y codificación de variables categóricas cuando fue necesario. Este proceso permitió asegurar la calidad de los datos antes del entrenamiento de los modelos.

Se implementaron cuatro algoritmos de clasificación supervisada: Naïve Bayes Gaussiano, Regresión Logística, Random Forest y Máquinas de Vectores de Soporte. Cada modelo posee hiperparámetros que influyen directamente en su capacidad de generalización y en el equilibrio entre sesgo y varianza, incluyendo el *trade-off* entre *overfitting* y *underfitting*.

En el caso de Naïve Bayes Gaussiano, se ajustó el parámetro de suavizado, que permite mejorar la estabilidad numérica y evitar problemas derivados de varianzas cercanas a cero. La Regresión Logística fue optimizada mediante el ajuste del parámetro de regularización, el tipo de penalización y el algoritmo de optimización, lo que permite controlar el sobreajuste y mejorar la convergencia. Para Random Forest se ajustaron el criterio de partición, la profundidad máxima de los árboles y la cantidad de variables consideradas en cada división, parámetros que afectan la complejidad del modelo y su capacidad de capturar relaciones no lineales. En el modelo SVM se optimizaron el parámetro de regularización, el tipo de kernel y el coeficiente gamma, que determinan la forma de la frontera de decisión.

La optimización de hiperparámetros se realizó mediante búsqueda en grilla combinada con validación cruzada *K-fold*, lo que permite evaluar el modelo en múltiples particiones del conjunto de datos y obtener estimaciones sólidas del desempeño.

Para la evaluación se utilizaron métricas estándar de clasificación, incluyendo *Accuracy*, *Precisión*, *Recall*, *F1-Score* y el Área Bajo la Curva ROC (AUC). Adicionalmente, se aplicó el mé-

todo de importancia por permutación, que consiste en medir la disminución del desempeño del modelo al alterar aleatoriamente cada variable, permitiendo estimar su contribución relativa.

RESULTADOS OBTENIDOS

Los resultados obtenidos en la Tabla 1 muestran que el modelo Random Forest presentó el mejor desempeño en ambos conjuntos de datos, seguido por SVM y Regresión Logística. Esto confirma la capacidad de los métodos de aprendizaje automático para capturar relaciones complejas en los datos clínicos.

Modelo	Caso	Accuracy	F1	AUC
Random Forest	Binario	0.8780	0.8775	0.9315
SVM (RBF)	Binario	0.8616	0.8609	0.9232
Random Forest	Multiclase	0.9432	0.9412	0.9868
Regresión Logística	Multiclase	0.8978	0.8953	0.9644

Cuadro 1: Resultados finales con las mejores configuraciones.

Los resultados obtenidos muestran que el modelo Random Forest presentó el mejor desempeño en ambos problemas, superando consistentemente a los demás algoritmos en las métricas evaluadas. Como se observa en las Figuras 1 y 2, este modelo alcanzó los valores más altos de Accuracy, F1-Score y AUC tanto en el caso binario como en el multiclase. El modelo SVM con kernel RBF también mostró un desempeño competitivo, especialmente en el caso binario, mientras que la Regresión Logística presentó resultados sólidos en el problema multiclase, aunque con menor capacidad para capturar relaciones no lineales complejas en comparación con los métodos de ensemble. Estos resultados confirman la efectividad de los modelos basados en múltiples árboles de decisión para el análisis de datos clínicos.

El análisis de importancia por permutación permitió evaluar cómo influye cada variable en el desempeño del modelo. Como se observa en las Figuras 3 y 4, el F1-Score aumenta progresivamente a medida que se incorporan los atributos en orden decreciente de importancia, lo que indica que las variables mejor posicionadas aportan

la mayor capacidad predictiva. Las primeras características generan el mayor incremento en el desempeño, mientras que las restantes contribuyen con mejoras marginales, evidenciando que una parte reducida de los atributos concentra la mayor parte de la información relevante. Este comportamiento confirma la coherencia del método de importancia por permutación y permite identificar las variables más influyentes en la clasificación.

El análisis de importancia por permutación permitió identificar las variables más influyentes en las predicciones. En el conjunto *HeartFailure*, la variable *ST_Slope* presentó la mayor contribución, mientras que en *Cardiotocography* la variable *ASTV* fue la más relevante.

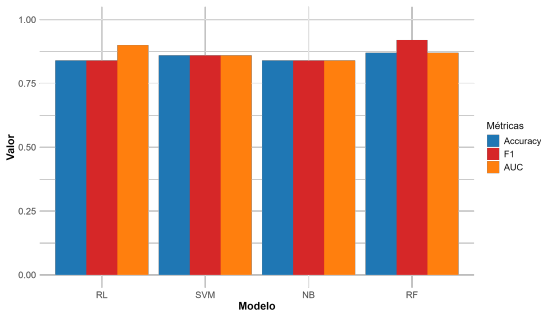


Figura 1: Comparación de desempeño entre modelos, caso binario.

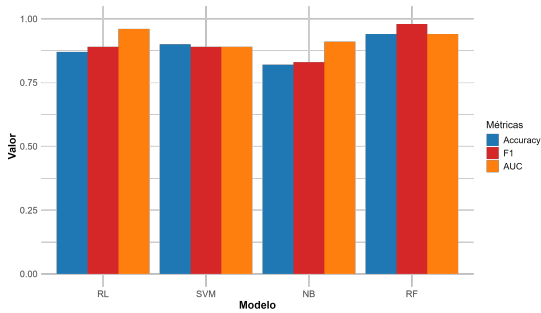


Figura 2: Comparación de desempeño entre modelos, caso multiclase.

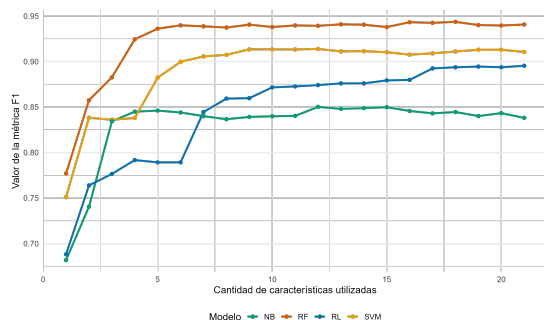


Figura 3: Importancia de variables mediante permutación para F1-Score, caso binario.

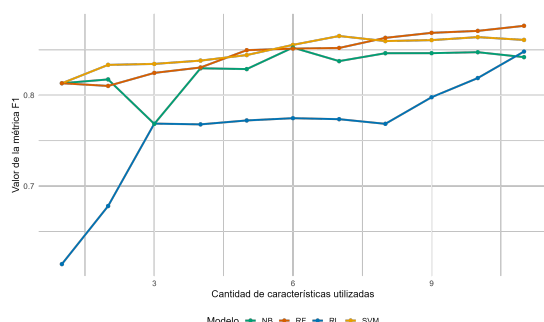


Figura 4: Importancia de variables mediante permutación para F1-Score, caso multiclase.

FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo del LIDEC se encuentra conformado por tres investigadores formados, cuatro investigadores estudiantes de doctorado, y cuatro alumnos de grado que se encuentran realizando la tesis final de grado. Este trabajo constituye la idea más importante de la tesis para obtener el título de Técnico Universitario en Inteligencia Artificial, del alumno Nicolás Seivane.

CONCLUSIONES

Se concluye que el modelo Random Forest es la herramienta más robusta para este tipo de datos debido a su capacidad para capturar relaciones no lineales y reducir el sobreajuste mediante el voto mayoritario de sus árboles. SVM

también mostró un rendimiento competitivo, especialmente con kernels RBF y polinómicos. El estudio permitió identificar atributos críticos como *ST_Slope* y *Oldpeak* para el diagnóstico de riesgo cardíaco. Como trabajos futuros, se propone el uso de optimización bayesiana para un ajuste de hiperparámetros más fino y técnicas de reducción de dimensionalidad (PCA) para simplificar los modelos sin perder precisión.

BIBLIOGRAFÍA

Referencias

- [1] D. J. Hand and K. Yu, “Idiot’s bayes—not so stupid after all?,” *International statistical review*, vol. 69, no. 3, pp. 385–398, 2001.
- [2] J. S. Cramer, “The origins of logistic regression,” tech. rep., Tinbergen Institute discussion paper, 2002.
- [3] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [4] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [5] J. L. Isaksen, M. Nørregaard, M. Manninger, D. Dobrev, T. Jespersen, B. Hermans, J. Heijman, G. Plank, D. Scherr, T. Pock, *et al.*, “Evaluating artificial intelligence-enabled medical tests in cardiology: Best practice,” *IJC Heart & Vasculture*, vol. 60, p. 101783, 2025.
- [6] N. Kumar and D. Kumar, “Machine learning based heart disease diagnosis using non-invasive methods: A review,” in *Journal of Physics: Conference Series*, vol. 1950, p. 012081, IOP Publishing, 2021.
- [7] fedesoriano, “Heart failure prediction dataset.” <https://www.kaggle.com/fedesoriano/heart-failure-prediction><https://www.kaggle.com/fedesoriano/heart-failure-prediction>, Sept. 2021.

- [8] J. Campos, D. y Bernar-
des, “Cardiotocography.” <https://doi.org/10.24432/C51S4N>, 2000.