

PRACTICA 1

KNIME

VICTOR GONZALEZ DEL CAMPO

vgd1005@alu.ubu.es

ANALISIS DE RESULTADOS DE EXPERIMENTOS:

DECISION TREE:

Row ID	T	F
T	2639	188
F	181	462

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
T	2639	181	462	188	0.933	0.936	0.933	0.719	0.935	?	?
F	462	188	2639	181	0.719	0.711	0.719	0.933	0.715	?	?
Overall	?	?	?	?	?	?	?	?	?	0.894	0.649

- **Verdaderos positivos (TP):** 2639. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "vivo" (T) y efectivamente pertenece a esa clase.
- **Falsos negativos (FN):** 188. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "muerto" (F) cuando en realidad pertenece a la clase "vivo" (T).
- **Falsos positivos (FP):** 181. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "vivo" (T) cuando en realidad pertenece a la clase "muerto" (F).

- **Verdaderos negativos (TN):** 462. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "muerto" (F) y efectivamente pertenece a esa clase.

La interpretación de la matriz de confusión nos permite evaluar el rendimiento del modelo. En este caso, el Decision Tree muestra un buen rendimiento en la clasificación de la clase "vivo" (T), ya que tiene un alto número de verdaderos positivos (2639) en comparación con los falsos negativos (188). Sin embargo, la cantidad de falsos positivos (181) indica que hay instancias clasificadas incorrectamente como "vivo" cuando en realidad son "muerto".

El accuracy del modelo es del 0.894, lo que indica que el 89.4% de las predicciones son correctas. El coeficiente de kappa de Cohen es de 0.694, que es una medida de acuerdo entre las clasificaciones observadas y las esperadas por el azar, donde 1 indica un acuerdo perfecto y 0 indica un acuerdo igual al azar.

Estos resultados muestran que el modelo Decision Tree tiene un buen rendimiento en general, pero también puede haber margen para mejorar su capacidad de distinguir entre las clases "vivo" y "muerto", especialmente reduciendo el número de falsos positivos.

FUZZY RULE:

Row ID	I T	I F
T	2591	112
F	247	336

Row ID	I TruePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Cohen's kappa
T	2591	247	336	112	0.959	0.913	0.959	0.576	0.935	?	?
F	336	112	2591	247	0.576	0.75	0.576	0.959	0.652	?	?
Overall	?	?	?	?	?	?	?	?	?	0.891	0.588

- **Verdaderos positivos (TP):** 2591. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "vivo" (T) y efectivamente pertenece a esa clase.

- **Falsos negativos (FN):** 112. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "muerto" (F) cuando en realidad pertenece a la clase "vivo" (T).
- **Falsos positivos (FP):** 247. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "vivo" (T) cuando en realidad pertenece a la clase "muerto" (F).
- **Verdaderos negativos (TN):** 336. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "muerto" (F) y efectivamente pertenece a esa clase.

El algoritmo Fuzzy Rule muestra un buen rendimiento en la clasificación de la clase "vivo" (T), con un alto número de verdaderos positivos (2591) en comparación con los falsos negativos (112). Sin embargo, la cantidad de falsos positivos (247) indica que hay instancias clasificadas incorrectamente como "vivo" cuando en realidad son "muerto".

El accuracy del modelo es del 0.891, lo que indica que el 89.1% de las predicciones son correctas. El coeficiente de kappa de Cohen es de 0.588, que es una medida de acuerdo entre las clasificaciones observadas y las esperadas por el azar, donde 1 indica un acuerdo perfecto y 0 indica un acuerdo igual al azar.

Estos resultados muestran que el modelo Fuzzy Rule tiene un buen rendimiento en general, pero puede haber margen para mejorar su capacidad de distinguir entre las clases "vivo" y "muerto", especialmente reduciendo el número de falsos positivos.

BAYES:

Row ID	I T	I F
T	47	2780
F	2	641

Row ID	I TruePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Cohen'...
T	47	2	641	2780	0.017	0.959	0.017	0.997	0.033	?	?
F	641	2780	47	2	0.997	0.187	0.997	0.017	0.315	?	?
Overall	?	?	?	?	?	?	?	?	?	0.198	0.005

- **Verdaderos positivos (TP):** 47. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "vivo" (T) y efectivamente pertenece a esa clase.

- Ahora, analicemos los resultados del algoritmo Naive Bayes:

- En resumen, los resultados del algoritmo Naive Bayes muestran un rendimiento extremadamente deficiente en la tarea de clasificación. La gran cantidad de falsos negativos indica que el modelo tiene dificultades para identificar correctamente las instancias de la clase "vivo". Además, el bajo valor de accuracy y coeficiente de kappa sugiere que el modelo no es confiable para realizar predicciones precisas. Es posible que se necesite revisar el conjunto de datos y la configuración del modelo para mejorar su rendimiento.

Row ID	<input type="checkbox"/> T	<input type="checkbox"/> F
T	2768	59
F	212	431

[illegible]

- **Verdaderos positivos (TP):** 2768. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "vivo" (T) y efectivamente pertenece a esa clase.
- **Falsos negativos (FN):** 59. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "muerto" (F) cuando en realidad pertenece a la clase "vivo" (T). Es importante destacar que estos casos son críticos, ya que representan situaciones en las que el modelo no logra detectar a personas que realmente necesitan atención médica.
- **Falsos positivos (FP):** 212. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "vivo" (T) cuando en realidad pertenece a la clase "muerto" (F). Aunque son errores, son menos críticos que los falsos negativos, ya que en este caso, el modelo está generando falsas alarmas.
- **Verdaderos negativos (TN):** 431. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "muerto" (F) y efectivamente pertenece a esa clase.

Ahora, consideremos los resultados del algoritmo Random Forest:

- El accuracy del modelo es del 0.992, lo que indica que el 99.2% de las predicciones son correctas. Esto es bastante alto y sugiere que el modelo tiene un excelente desempeño en general.
- El coeficiente de kappa de Cohen es de 0.715, lo cual es bastante bueno. Este valor indica un acuerdo sustancial entre las clasificaciones observadas y las esperadas por el azar.

Es crucial destacar que, en un contexto médico como este, es fundamental minimizar los falsos negativos. Es decir, es más importante identificar correctamente a las personas que están en riesgo de fallecer y asegurarse de que reciban la atención necesaria a tiempo. Aunque es bueno tener un alto nivel de precisión global (accuracy), el enfoque principal debería estar en reducir los falsos negativos, incluso si esto implica un ligero aumento en los falsos positivos.

En resumen, los resultados del algoritmo Random Forest son muy prometedores, con un alto nivel de precisión y un coeficiente de kappa considerable. Sin embargo, se debe prestar especial atención a la reducción de los falsos negativos para mejorar aún más la capacidad del modelo para identificar a las personas en riesgo.

GRADIENT BOOSTER:

Row ID	I T	I F
T	2762	65
F	186	457

Row ID	I TruePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Cohen'...
T	2762	186	457	65	0.977	0.937	0.977	0.711	0.957	?	?
F	457	65	2762	186	0.711	0.875	0.711	0.977	0.785	?	?
Overall	?	?	?	?	?	?	?	?	?	0.928	0.742

- **Verdaderos positivos (TP):** 2762. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "vivo" (T) y efectivamente pertenece a esa clase.
- **Falsos negativos (FN):** 65. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "muerto" (F) cuando en realidad pertenece a la clase "vivo" (T). Como mencionamos anteriormente, los falsos negativos son críticos en un contexto médico, ya que representan situaciones en las que el modelo no logra detectar a personas que necesitan atención urgente.
- **Falsos positivos (FP):** 186. Estos son los casos en los que el modelo predijo incorrectamente que la instancia pertenece a la clase "vivo" (T) cuando en realidad pertenece a la clase "muerto" (F). Aunque son errores, son menos críticos que los falsos negativos, ya que representan falsas alarmas.
- **Verdaderos negativos (TN):** 457. Estos son los casos en los que el modelo predijo correctamente que la instancia pertenece a la clase "muerto" (F) y efectivamente pertenece a esa clase.

Ahora, analicemos los resultados del algoritmo Gradient Booster:

- El accuracy del modelo es del 0.928, lo que indica que el 92.8% de las predicciones son correctas. Este es un buen nivel de precisión en general.
- El coeficiente de kappa de Cohen es de 0.742, lo cual es bastante alto. Este valor indica un buen acuerdo entre las clasificaciones observadas y las esperadas por el azar.

En conclusión, los resultados del algoritmo Gradient Booster son muy alentadores, con un alto nivel de precisión y un coeficiente de kappa significativo. Sin embargo, al igual que con otros algoritmos, es fundamental reducir los falsos negativos para garantizar que las personas que están en riesgo reciban la atención médica necesaria a tiempo. Esto puede ayudar a mejorar aún más la capacidad del modelo para identificar a las personas en situación de riesgo.

ANALISIS GENERAL:

Después de analizar detalladamente los resultados de los cinco algoritmos de clasificación (Decision Tree, Fuzzy Rule, Bayes, Random Forest y Gradient Booster), podemos extraer las siguientes conclusiones:

1. **Random Forest destaca como el mejor algoritmo:** Con un impresionante accuracy del 99.2% y un coeficiente de Cohen's Kappa de 0.715, el Random Forest demuestra una capacidad excepcional para realizar predicciones precisas. Su capacidad para identificar correctamente los casos de "vivo" y "muerto", junto con un bajo número de falsos negativos, lo convierte en una opción sobresaliente para este tipo de análisis.
2. **Gradient Booster muestra un rendimiento sólido:** Con un accuracy del 92.8% y un coeficiente de Cohen's Kappa de 0.742, el Gradient Booster se sitúa como una alternativa confiable para la clasificación de datos. Aunque no alcanza el nivel de precisión del Random Forest, su capacidad para identificar correctamente los casos de interés (en este caso, las personas que podrían fallecer) es notable.
3. **Decision Tree y Fuzzy Rule ofrecen resultados aceptables:** Ambos algoritmos muestran un rendimiento decente, con accuracies por encima del 89% y coeficientes de Cohen's Kappa superiores a 0.5. Sin embargo, su capacidad para distinguir entre las clases "vivo" y "muerto" podría ser mejorada, especialmente en la reducción de falsos positivos.
4. **Bayes presenta el rendimiento más bajo:** Con un accuracy del 18.9% y un coeficiente de Cohen's Kappa cercano a cero, el Bayes muestra un rendimiento insatisfactorio para este conjunto de datos. Su capacidad para predecir correctamente las clases de interés es muy limitada, lo que sugiere que no es la mejor opción para este escenario específico.

En resumen, el Random Forest emerge como el mejor algoritmo en base a los resultados obtenidos, gracias a su capacidad excepcional para realizar predicciones precisas y su baja tasa de falsos negativos. Sin embargo, el Gradient Booster también muestra un rendimiento destacable y podría ser considerado como una alternativa viable dependiendo de las necesidades y restricciones del proyecto.

