

Predicción de Enfermedades Cardíacas

Proyecto 2

Analítica Computacional para la Toma de Decisiones

Universidad de los Andes, Bogotá, Colombia

Santiago González Montealegre	s.gonzalez35	202012274
Juliana Carolina Cárdenas Barragán	jc.cardenasb1	202011683

Fecha de presentación: Mayo 4 de 2023

Tabla de contenido

1	Introducción	1
2	Modelos.....	1
2.1	Modelo Basado en Investigación	2
2.2	Modelo Basado en otros Autores	2
3	Training	2
4	Testing.....	3
4.1	Método de Clasificación	3
5	Evaluación de Modelos	3
5.1	Medidas de Desempeño	3
5.2	Rendimiento Modelos	5
5.2.1	Modelo Original	5
5.2.2	Aprendizaje de Estructura: Método de Restricciones	5
5.2.3	Aprendizaje de Estructura: Puntaje K2	5
5.2.4	Aprendizaje de Estructura: Puntaje BIC	6
5.2.5	Modelo Basado en otros Autores	6
5.3	Selección de Modelo	6

1 Introducción

Este reporte muestra el proceso de creación, entrenamiento, testeo y evaluación de diferentes modelos de Redes Bayesianas para predecir la presencia de una enfermedad cardíaca en un paciente dados ciertos inputs. Para esta tarea se trabajó con un data set de pacientes del centro médico *Long Beach and Cleveland Clinic Foundation*. Este cuenta con 303 registros y cada uno de estos registros tiene información de 14 atributos demográficos o médicos del paciente.

2 Modelos

Haciendo uso de estos datos se desarrollaron distintos modelos con el fin de predecir o evaluar si un paciente tiene o no una Enfermedad Cardíaca, es decir, con base en 13 atributos usados como variables construir los modelos que permiten predecir el valor del target (Enfermedad Cardíaca). En total se desarrollaron 5 modelos: modelo con base en investigación (iteración 1 del proyecto), modelo basado en aprendizaje de estructura con restricciones, modelo basado en aprendizaje de estructura con puntaje K2, modelo basado en aprendizaje de

estructura con puntaje BIC y un modelo basado en otros autores. El objetivo es evaluar los distintos modelos con el fin de poder ordenarlos según su utilidad para el problema en cuestión.

2.1 Modelo Basado en Investigación

Tras realizar un modelo inicial basado en estadística descriptiva e información de autores respetados dentro del campo de estudio, se corrigieron ciertos aspectos y se obtuvo finalmente el siguiente modelo:

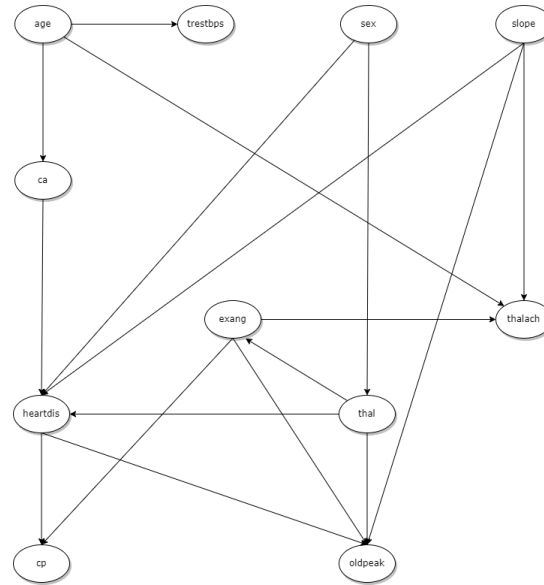


Figura 1. Modelo Basado Investigación.

2.2 Modelo Basado en otros Autores

El modelo propuesto por otros autores que tienen el mismo objeto de estudio se ve de la siguiente forma:

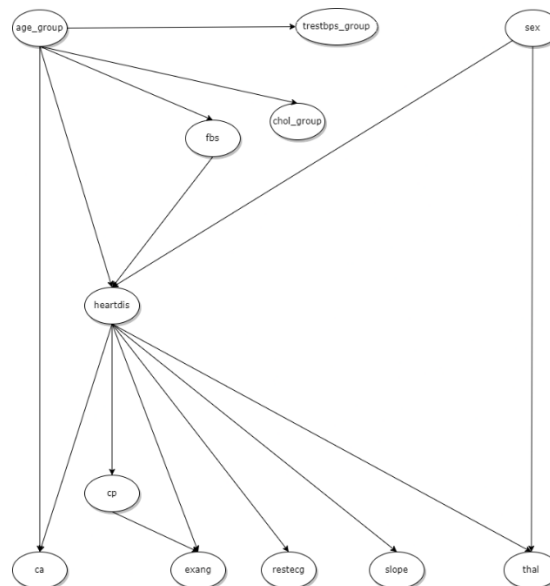


Figura 2. Modelo Basado en otros Autores.

3 Training

Para el proceso de entrenamiento de los modelos es necesario ser cautelosos con la metodología, ya que es necesario que se entrenen todos los modelos con los mismos datos y posteriormente sean probados con los mismos datos. Para esta tarea generalmente se utiliza la

Rule of Thumb de entrenar el modelo con el 75% de los datos y probarlo con el 25% restante, generalmente seleccionados aleatoriamente. Sin embargo, esta metodología genera problemas en data sets con pocos registros ya que por aleatoriedad se puede terminar en el set de prueba con datos de muy difícil o muy fácil clasificación. En consecuencia, las métricas o medidas de desempeño obtenidas con este método no son confiables.

Con el fin de evitar este problema ya que contamos con únicamente 303 datos, usamos el método conocido como *k-fold Cross Validation*. Este método consiste básicamente en dividir el conjunto de datos en k subconjuntos, posteriormente se realizan k iteraciones donde se selecciona el subconjunto i como el test set, mientras que los demás $k - 1$ subconjuntos se usan como el training set. Con esto eliminamos el posible error en las medidas por la aleatoriedad, ya que calculamos para cada iteración las medidas de desempeño y finalmente se promedian. Para el entrenamiento de los modelos se usó $k = 4$ y el proceso se ve de la siguiente forma:

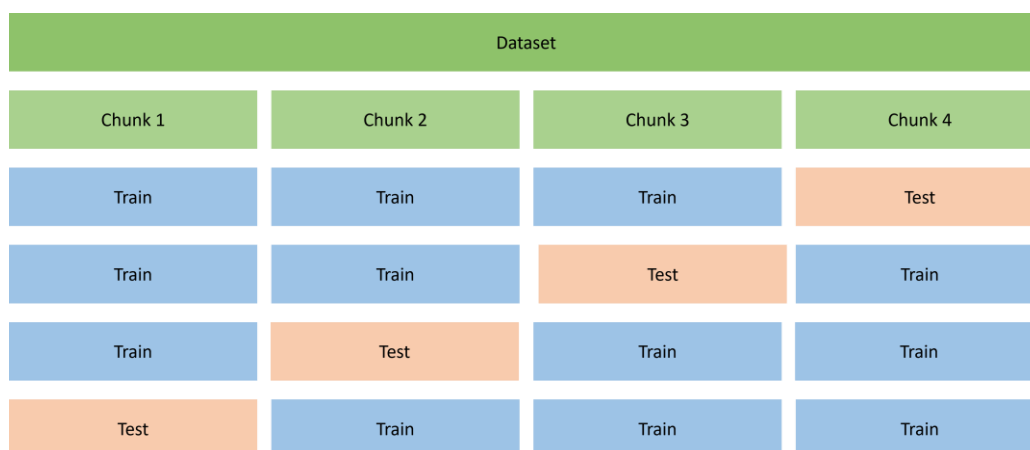


Figura 3. K-fold Cross Validation

4 Testing

4.1 Método de Clasificación

El método usado para clasificar si el paciente tiene enfermedad cardíaca o no consiste en obtener las probabilidades dada la evidencia, los datos de la persona que se usan como variables en la red bayesiana. Por lo tanto, cuando la probabilidad de que tenga enfermedad cardíaca sea mayor a la probabilidad de que no tenga enfermedad cardíaca se clasifica a la persona como que tiene enfermedad cardíaca, de lo contrario, se clasifica como no tiene enfermedad cardíaca. Este será el método usado para probar cada uno de los modelos en la etapa de evaluación de modelos con el test set que se obtuvo con el método *k-fold Cross Validation*.

5 Evaluación de Modelos

5.1 Medidas de Desempeño

Se definieron 5 medidas de desempeño con el fin de probar que tanto del problema real logra solucionar el modelo. Para entender estas medidas primero es necesario definir la matriz de confusión que nos permite separar en conjuntos los datos de la clasificación, a continuación, se muestra un ejemplo:

		Actual	
		No Heart Disease	Heart Disease
Predicted	No Heart Disease	TN	FN
	Heart Disease	FP	TP

Figura 4. Matriz de Confusión

Esta matriz se interpreta de la siguiente forma:

- TN: True Negatives, representa la cuenta de pacientes que no tenían enfermedad cardíaca y fueron clasificados o predichos como que no tenían enfermedad cardíaca. Es decir, una clasificación negativa correcta.
- FN: False Negatives, representa la cuenta de pacientes que tenían enfermedad cardíaca y fueron clasificados o predichos como que no tenían enfermedad cardíaca. Es decir, una clasificación negativa incorrecta.
- FP: False Positives, representa la cuenta de pacientes que no tenían enfermedad cardíaca y fueron clasificados o predichos como que tenían enfermedad cardíaca. Es decir, una clasificación positiva incorrecta.
- TP: True Positives, representa la cuenta de pacientes que tenían enfermedad cardíaca y fueron clasificados o predichos como que tenían enfermedad cardíaca. Es decir, una clasificación positiva correcta.

Se desea únicamente obtener clasificaciones correctas con nuestro modelo, sin embargo, esto es poco realista ya que no existe un modelo que represente a la perfección la situación real. Por lo tanto, se desea maximizar el valor de TN y TP o minimizar el valor de FN y FP.

Sin embargo, observar cuentas de cada valor dificulta la interpretación, por lo que es mejor tener ciertas medidas estandarizadas que permitan evaluar un modelo y de hecho compararlo con otros. Estas medidas se basan en los 4 elementos de la matriz y son las siguientes:

- Accuracy: Es el porcentaje de predicciones que fueron correctas.

$$Accuracy = \frac{TN + TP}{TN + FN + FP + TP}$$

Sin embargo, esta medida puede ser sesgada por lo que se definen otras medidas de desempeño.

- Precision: Este valor representa que tan preciso es el modelo con las predicciones positivas que realiza.

$$Precision = \frac{TP}{FP + TP}$$

- Recall: Este valor representa la proporción de casos positivos que el modelo predice correctamente

$$Recall = \frac{TP}{FN + TP}$$

La idea del modelo es predecir si un paciente tiene o no una enfermedad cardíaca, por lo que predecir como que no tiene la enfermedad cuando en realidad si tiene la enfermedad es

bastante grave ya que es la vida de una persona. Por otro lado, predecir que el paciente tiene la enfermedad cardíaca cuando en realidad no la tiene, supone un problema menor, quizás se le hagan más estudios al paciente pero que posteriormente descartarán el diagnóstico.

En resumen, en nuestro caso, no en general, es mejor tener menos FN, a cambio de obtener mas FP. Esto debido a que tienen una relación inversamente proporcional entre ambos. Es decir, se prioriza tener un mayor *recall* a cambio de perder *precision*, este elemento se usará en la evaluación de modelos.

- F1 Score: Creamos una métrica que combine en un solo valor tanto Precision como Recall, para esto se usa la media armónica entre ambos valores, la cual le otorga un mayor peso a los valores pequeños.

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

- Specificity: Este valor representa la proporción de casos negativos que el modelo predice correctamente

$$Specificity = \frac{TN}{TN + FP}$$

5.2 Rendimiento Modelos

Ahora que son claras las medidas a usar para evaluar los modelos, se procede a calcularlas para cada modelo con la metodología de k-fold Cross Validation. Esto consiste en calcular las 5 medidas de desempeño definidas anteriormente para los 4 test set definidos y finalmente se calcula la media de estos valores para cada medida de desempeño para obtener el promedio del modelo.

5.2.1 Modelo Original

Con el modelo original se obtuvieron los siguientes resultados:

TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Specificity
27	14	5	30	0.75	0.68	0.86	0.76	0.66
39	6	11	20	0.78	0.77	0.65	0.70	0.87
29	8	8	31	0.79	0.79	0.79	0.79	0.78
31	10	10	24	0.73	0.71	0.71	0.71	0.76
Promedio				0.76	0.74	0.75	0.74	0.77

Figura 5. K-fold Cross Validation Original

5.2.2 Aprendizaje de Estructura: Método de Restricciones

Con el modelo que usa aprendizaje de estructura con el método de restricciones se obtuvieron los siguientes resultados:

TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Specificity
26	15	8	27	0.70	0.64	0.77	0.70	0.63
40	5	10	21	0.80	0.81	0.68	0.74	0.89
29	8	10	29	0.76	0.78	0.74	0.76	0.78
37	4	17	17	0.72	0.81	0.50	0.62	0.90
Promedio				0.75	0.76	0.67	0.70	0.80

Figura 6. K-fold Cross Validation Restricciones

5.2.3 Aprendizaje de Estructura: Puntaje K2

Con el modelo que usa aprendizaje de estructura con el método de puntaje K2 se obtuvieron los siguientes resultados:

TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Specificity
32	9	6	29	0.80	0.76	0.83	0.79	0.78
33	12	6	25	0.76	0.68	0.81	0.74	0.73
30	7	10	29	0.78	0.81	0.74	0.77	0.81
32	9	6	28	0.80	0.76	0.82	0.79	0.78
Promedio				0.79	0.75	0.80	0.77	0.78

Figura 7. K-fold Cross Validation K2 Score

5.2.4 Aprendizaje de Estructura: Puntaje BIC

Con el modelo que usa aprendizaje de estructura con el método de puntaje K2 se obtuvieron los siguientes resultados:

TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Specificity
35	6	7	28	0.83	0.82	0.80	0.81	0.85
40	5	6	25	0.86	0.83	0.81	0.82	0.89
33	4	10	29	0.82	0.88	0.74	0.81	0.89
38	3	6	28	0.88	0.90	0.82	0.86	0.93
Promedio				0.84	0.86	0.79	0.82	0.89

Figura 8. K-fold Cross Validation BIC Score

5.2.5 Modelo Basado en otros Autores

Con el modelo que usa aprendizaje de estructura con el método de puntaje K2 se obtuvieron los siguientes resultados:

TN	FP	FN	TP	Accuracy	Precision	Recall	F1 Score	Specificity
34	7	8	26	0.80	0.79	0.76	0.78	0.83
34	7	8	25	0.80	0.78	0.76	0.77	0.83
33	4	6	31	0.86	0.89	0.84	0.86	0.89
35	6	7	26	0.82	0.81	0.79	0.80	0.85
Promedio				0.82	0.82	0.79	0.80	0.85

Figura 8. K-fold Cross Validation BIC Score

5.3 Selección de Modelo

Se resumen los resultados de todos los modelos en la siguiente tabla:

	Accuracy	Precision	Recall	F1 Score	Specificity
Original	0.76	0.74	0.75	0.74	0.77
Restricciones	0.75	0.76	0.67	0.7	0.8
K2 Score	0.79	0.75	0.8	0.77	0.78
BIC Score	0.84	0.86	0.79	0.82	0.89
Others	0.82	0.82	0.79	0.8	0.85

Figura 9. Rendimiento Modelos

Como se definió anteriormente nuestra prioridad se encuentra en el Recall, ya que deseamos tener el menor número de falsos negativos. Ante esto observamos que el mejor modelo es el de aprendizaje de estructura por medio del puntaje K2. Sin embargo, este tiene menor precision que otros modelos por lo que analizando el F1 Score que combina ambas medidas observamos que el modelo con BIC Score y el modelo de otros autores lo superan y tienen un Recall muy similar.

Para decidir de mejor forma se hará un gráfico de la curva ROC, la cual consiste en graficar la Specificity contra el Recall con el fin de analizar el mejor modelo. Estas dos medidas se eligen ya que el Recall y la Precision están relacionados y poseen información similar, si nos

enfocamos únicamente en estos se dejaría de lado los verdaderos negativos y su peso en el modelo. A continuación, se muestra el gráfico:

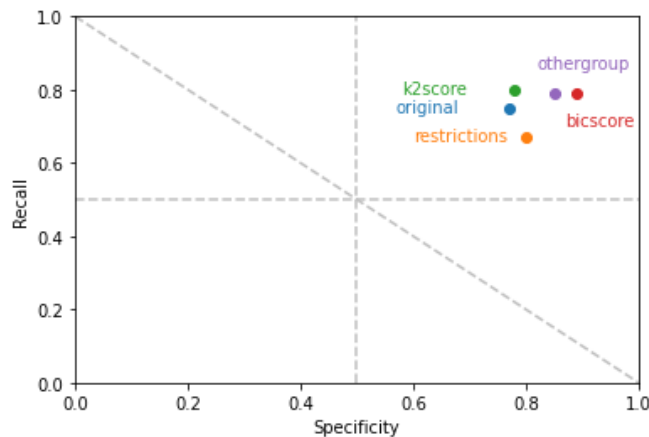


Figura 10. ROC Curve

La línea diagonal que divide el gráfico representa la línea límite de un modelo general, ya que esta línea representa un modelo que clasifica de manera aleatoria si tiene enfermedad cardíaca o no. Por lo que cualquier modelo que se ubique por debajo de esta línea será descartado automáticamente ya que es peor que un modelo aleatorio.

Podemos ver que todos los modelos construidos son mejores que un modelo aleatorio, por lo que hasta el peor de ellos proporciona valor en el proceso de clasificación. Sin embargo, se desea seleccionar el mejor modelo. Por lo que iniciamos descartando los modelos Original y de Restricciones ya que hay por lo menos un modelo que los supera en todas las medidas de desempeño.

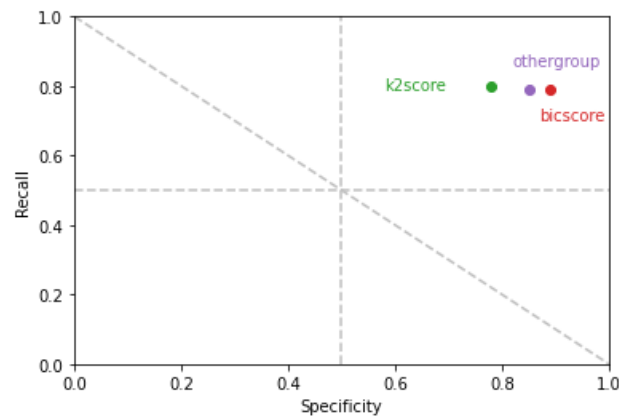


Figura 11. ROC Curve Filtered

El modelo K2 Score aporta un mayor Recall contra los otros modelos, 0.80 vs 0.79. Pero, se comporta bastante peor frente a la Specificity de los demás modelos, por lo que se decide perder un 1% de Recall a cambio de mejorar considerablemente la Specificity, descartamos el modelo K2 Score. Entre el modelo con BIC Score y el de otros autores tienen el mismo Recall, pero el BIC Score tiene mejor Specificity y mejor F1 Score, por lo que el mejor modelo que podemos encontrar es el de aprendizaje de estructura con BIC Score.