

Caso de estudio

Enfermedades Cardiológicas

Introducción a los métodos de aprendizaje
automático

Septiembre 2019



Universidad
Católica del
Uruguay

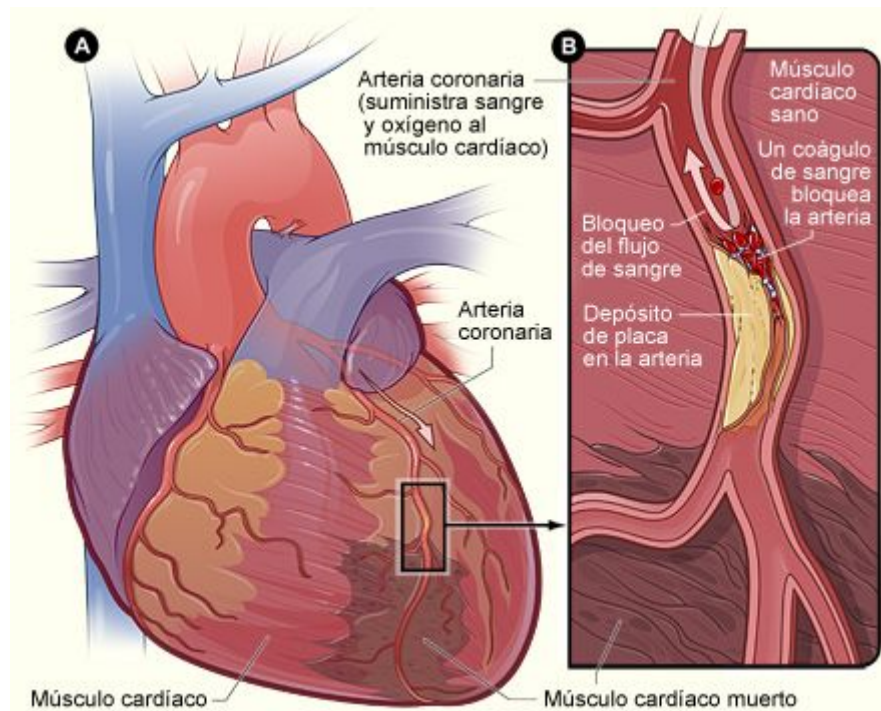
Autores:

- Santiago Casás
- Agustín Betancor
- Martín Rose
- Mauricio Coniglio
- Gianni Iaquina

Índice

Problema	3
Función del sistema circulatorio	4
Funcionamiento del corazón	5
Conjuntos de Datos	6
Análisis	10
Integración	10
Determinación de los atributos más relevantes y sus estadísticas	10
Datos de interés	10
Listado de atributos relevantes para la solución	10
Análisis estadístico	12
Outliers	12
Reemplazo de Missings	13
Selección del Método de Aprendizaje	13
Diagrama del Proceso en RapidMiner	14
Conjunto de Entrenamiento y Testing	14
Análisis de los Resultados	14
Referencias	15

Problema



<http://www.monografias.com/trabajos95/el-sistema-circulatorio/image009.jpg>

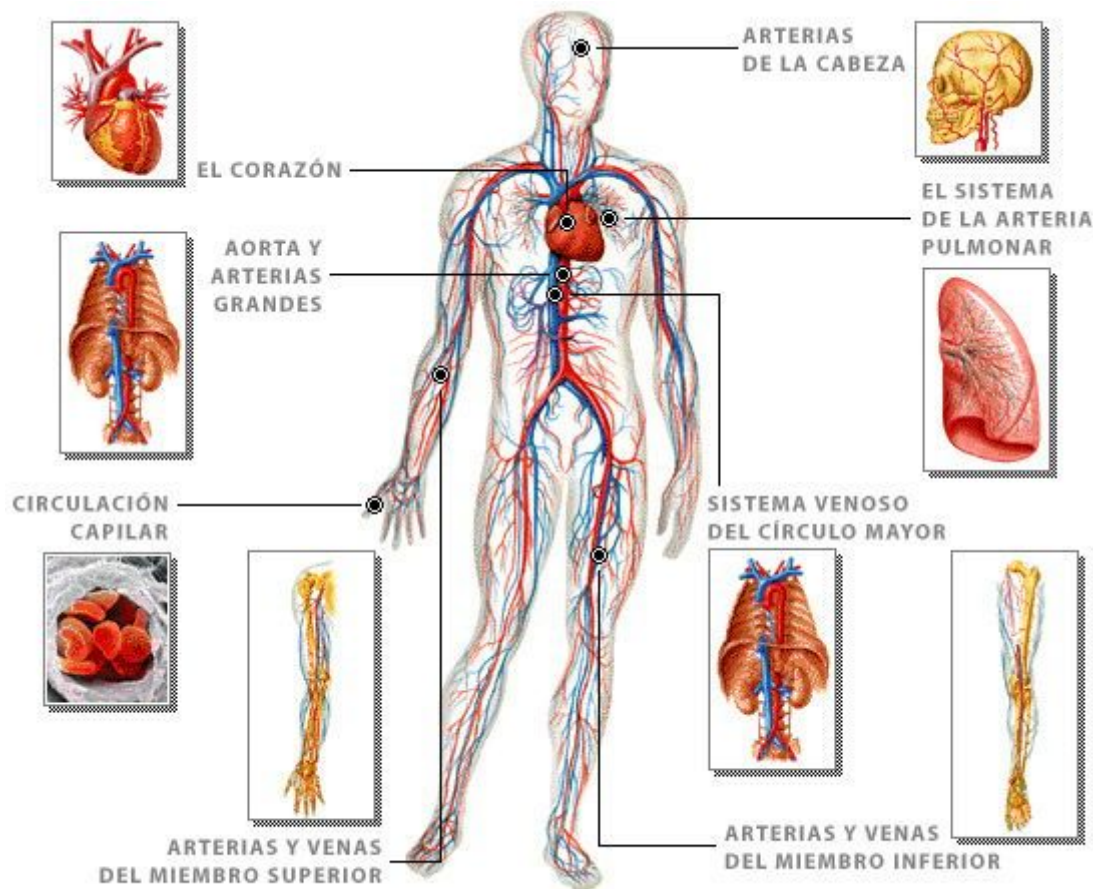
El problema se basa en el estudio de enfermedades cardíacas, a partir del estudio de 4 dataset de datos tomados de 4 lugares geográficamente distantes (Cleveland - EE.UU. EAST, Hungría, Suiza, y VA Long Beach - EE.UU. WEST), obtenidos del repositorio de Machine Learning, del sitio web de la Universidad de California (Irvine). El objetivo principal de este estudio es la predicción de la existencia de enfermedades al corazón del paciente. Estos 4 datasets tienen el mismo formato de datos, con los mismos atributos. La idea de que los datos sean geográficamente distantes, es poder generalizar no sólo para una región en particular, sino que abstraerse de la ubicación geográfica. A partir de estos datasets, se pretende lograr una predicción de si el paciente padece/rá (o no) una enfermedad cardíaca. Dado lo costoso de los tratamientos médicos para enfermedades cardíacas, éstos 4 datasets contienen información acerca de pacientes geográficamente dispersos y la intención es poder a partir de estos datos, predecir los pacientes con mayor riesgo de enfermedades cardíacas, y así tratar a este tipo de pacientes de forma preventiva, y no reactiva. Con esto disminuirían tanto costo de tratamiento reactivo, así como evitar posibles cuadros cardíacos.

En general, en los papers que han utilizado estos dataset como fuente para sus investigaciones, en la mayoría de los casos han utilizado sólo 14 de los 76 atributos de dichos datasets. En nuestro caso creemos que es relevante el uso de 20 de los 76 atributos, ya que algunos de los atributos influyen indirectamente en la probabilidad de diagnosticar una probable enfermedad cardíaca.

En muchos casos, si bien el atributo por sí sólo no influye significativamente en la probabilidad de tener enfermedades cardíacas, cuando son evaluados con algunos atributos más (en conjunto) aumenta significativamente la probabilidad. Por ejemplo, en el caso de

tener una persona de 50 años, que no hace deporte, fuma, y su ritmo cardíaco en reposo es elevado, ésta persona tiene altas probabilidades de tener una enfermedad cardíaca a futuro.

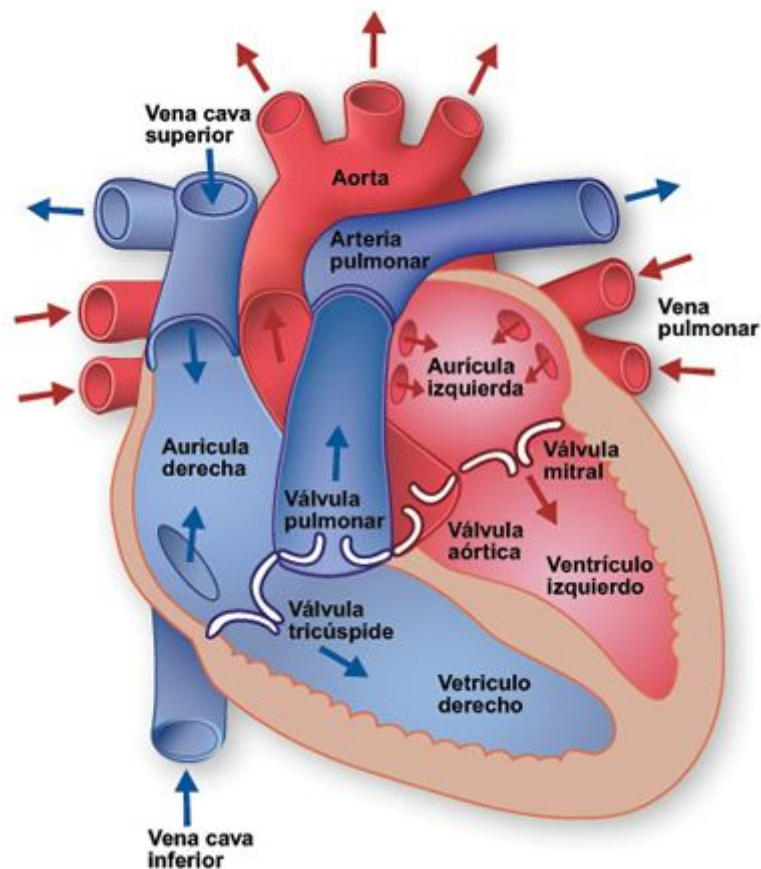
Función del sistema circulatorio



https://www.portaleducativo.net/biblioteca/sistema_circulatorio.jpg

La función principal del Sistema Circulatorio es enviar nutrientes, gases (O_2), hormonas y células sanguíneas al resto de las células del cuerpo. También recoge los desperdicios del metabolismo de las células del cuerpo (por ejemplo CO_2). El Sistema Circulatorio se compone del corazón, vasos sanguíneos (arterias, venas, capilares) y la sangre. Los órganos encargados de eliminar estos desechos son los riñones (mediante la orina), y los pulmones (al exhalar CO_2).

Funcionamiento del corazón



https://www.portaleducativo.net/biblioteca/corazon_1.jpg

Para entender mejor el problema, a continuación se detalla una breve introducción a el funcionamiento del corazón:

1. El corazón recibe sangre desoxigenada a través de las venas cavas (superior e inferior), esta sangre entra en el aurículo derecho.
2. El aurículo derecho se comprime, pasando la sangre (a través de la válvula tricúspide) al ventrículo derecho.
3. El ventrículo derecho se comprime, enviando la sangre en el mismo hacia la arteria pulmonar (a través de la válvula pulmonar).
4. La sangre se oxigena en los pulmones, para luego ingresar nuevamente al corazón a través de la vena pulmonar, ingresando en el aurículo izquierdo.
5. El aurículo izquierdo se comprime, mandando la sangre (a través de la válvula mitral) al ventrículo izquierdo (sangre oxigenada y con nutrientes).
6. El ventrículo izquierdo se comprime, y envía la sangre oxigenada (a través de la válvula aórtica) hacia la Arteria aorta, para que de esta manera se envíe sangre oxigenada al resto del cuerpo.

Conjuntos de Datos

Como se menciona en la sección anterior, los *data sets* están compuestos de 76 atributos. A continuación se listan los mismos junto con el tipo de dato que contienen y su descripción.

#	Nombre	Descripción	Tipo de dato
1	id	Número identificador del paciente	Integer
2	ccf	Número de seguridad social del paciente	Integer
3	age	Años de edad	Integer
4	sex	Sexo	Integer (1 = male, 0 = female)
5	painloc	Ubicación del dolor en el pecho	Integer (1 = substernal, 0 = otherwise)
6	painexer	Dolor provocado por esfuerzo	Integer (1 = provoked by exertion, 0 = otherwise)
7	relrest	Dolor aliviado luego de descansar	Integer (1 = relieved after rest, 0 = otherwise)
8	pncanden	Suma de los atributos 5, 6 y 7	Integer
9	cp	Tipo de dolor en el pecho	Integer (1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
10	trestbps	Presión sanguínea en reposo. Medida en mmHg al ingresar al hospital	Integer
11	htn	Paciente tiene hipertensión	Integer (1 = hypertension, 0 = no hypertension)
12	chol	Colesterol sérico. Medido en mg/dl	
13	smoke	Si el paciente es fumador	Integer (1 = yes, 0 = no)
14	cigs	Cigarros por día	Integer
15	years	Número de años de fumador	Integer
16	fbs	Indica si el azúcar en sangre en ayunas es mayor a 120 mg/dl	Integer (1 = true, 0 = false)
17	dm	Antecedentes de diabetes	Integer (1 = history of diabetes, 0 = no such history)
18	famhist	Antecedentes de enfermedad coronaria arterial en la familia	Integer (1 = yes, 0 = no)
19	restecg	Resultados electrocardiográficos en reposo	Integer (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria

20	ekgmo	Mes de la lectura del ejercicio del Electrocardiograma	Integer
21	ekgday	Día de la lectura del ejercicio del Electrocardiograma	Integer
22	ekgyr	Año de la lectura del ejercicio del Electrocardiograma	Integer
23	dig	Indica si se utilizó digitalis durante el Electrocardiograma	Integer (1 = yes, 0 = no)
24	prop	Indica si se utilizó algún betabloqueante durante el Electrocardiograma	Integer (1 = yes, 0 = no)
25	nitr	Indica si se utilizó algún nitrato durante el Electrocardiograma	Integer (1 = yes, 0 = no)
26	pro	Indica si se utilizó algún bloqueador de los canales de calcio durante el Electrocardiograma	Integer (1 = yes, 0 = no)
27	diuretic	Indica si se utilizó algún diurético durante el Electrocardiograma	Integer (1 = yes, 0 = no)
28	proto	Protocolo utilizado para el Electrocardiograma	Integer (1 = Bruce, 2 = Kottus, 3 = McHenry, 4 = fast Balke, 5 = Balke 6 = Noughton, 7 = bike 150 kpa min/min, 8 = bike 125 kpa min/min, 9 = bike 100 kpa min/min, 10 = bike 75 kpa min/min, 11 = bike 50 kpa min/min, 12 = arm ergometer)
29	thaldur	Duración del test del ejercicio (en minutos)	Integer
30	thalttime	Instante en el que se observó una depresión en la medida del segmento ST	Integer
31	met	Resultado del ejercicio en mets (Metabolic equivalents)	Integer
32	thalach	Mayor ritmo cardíaco observado	Integer
33	thalrest	Ritmo cardíaco en reposo	Integer
34	tpeakbps	Pico de la presión arterial en el ejercicio en la primera parte	Integer
35	tpeakbpd	Pico de la presión arterial en el ejercicio en la segunda parte	Integer
36	dummy	Indefinido	-
37	trestbpd	Presión sanguínea en reposo	Integer
38	exang	Angina inducida por el ejercicio	Integer (1 = yes, 0 = no)
39	xhypo	No se encontró información certera	Integer (1 = yes, 0 = no)
40	oldpeak	Depresión del segmento ST inducida por el ejercicio en relación con al reposo.	Integer
41	slope	Pendiente del segmento ST del ejercicio máximo	Integer (1 = upsloping, 2 = flat, 3 = downsloping)

42	rldv5	Altura en reposo	Integer
43	rldv5e	Altura en el ejercicio máximo	Integer
44	ca	Número de vasos principales coloreados por fluoroscopia	Integer
45	restckm	Irrelevante	-
46	exerckm	Irrelevante	-
47	restef	Fracción de eyección en reposo	Integer
48	restwm	Movimiento anormal en el movimiento de las paredes del corazón en reposo	Integer (0 = none, 1 = mild or moderate, 2 = moderate or severe, 3 = akinesis or dyskmem)
49	exeref	Fracción de eyección en el ejercicio	Integer
50	exerwm	Movimiento del corazón en el reposo	Integer
51	thal	Prueba de esfuerzo con talio	Integer (3 = normal, 6 = fixed defect, 7 = reversable defect)
52	thalsev	No se usa	-
53	thalpul	No se usa	-
54	earlobe	No se usa	-
55	cmo	Mes de la Cateterización cardiaca	Integer
56	cday	Día de la Cateterización cardiaca	Integer
57	cyr	Año de la Cateterización cardiaca	Integer
58	num	Diagnóstico de enfermedad del corazón (estado de la angiografía)	Integer (0 = < 50% diameter narrowing, 1 = > 50% diameter narrowing)
59	lmt	Indefinido	-
60	ladprox	Indefinido	-
61	laddist	Indefinido	-
62	diag	Indefinido	-
63	cxmain	Indefinido	-
64	ramus	Indefinido	-
65	om1	Indefinido	-
66	om2	Indefinido	-
67	rcaprox	Indefinido	-
68	rcadist	Indefinido	-
69	lvx1	No se usa	-

70	lvx2	No se usa	-
71	lvx3	No se usa	-
72	lvx4	No se usa	-
73	lvf	No se usa	-
74	cathef	No se usa	-
75	junk	No se usa	-
76	name	Nombre del paciente (reemplazado por "name" en los <i>data sets</i>)	String

Análisis

Las 4 bases de datos poseen los mismos atributos y estos son del mismo tipo por lo que la integración de las mismas se puede realizar sin ningún inconveniente. Por eso luego de detallar los atributos de las bases de datos pasamos a analizar cómo integrar las bases, luego determinar cuáles son los atributos más relevantes e independientes y como vamos a trabajar los datos faltantes para poder realizar un mejor trabajo de predicción.

Integración

Buscando alternativas para realizar el trabajo de integración de las bases pensamos en realizar un solo dataset con todas las instancias de todas las bases pero luego se encontró que el RapidMiner (RM) posee un operador "Append" que construye un dataset combinando dos o más dataset compatibles, donde añade todas las instancias de todos los dataset seleccionados.

También se utilizaron otros operadores para poder realizar el proceso en el RM y que de esta manera queden integrados todos los datos y poder realizar las predicciones sobre ellos.

Estos operadores son detallados luego, pero hacemos especial mención a los operadores "Join" y "Generate ID". Estos operadores se utilizan en el proceso de detección de outliers para poder separar individualmente las columnas que quieren ser analizadas del resto del dataset. Particularmente se utilizan para volver a unir las columnas al resto del datasets. El ID es necesario para mantener la relación de las tuplas.

Determinación de los atributos más relevantes y sus estadísticas

Datos de interés

Total de datos: 899

Total de atributos: 76

Cantidad de atributos utilizados por otros: 14

Cantidad de atributos relevantes empleados en la solución: 22

Listado de atributos relevantes para la solución

Si bien en la mayoría de los papers en los cuales se hacía referencia a este dataset utilizaban para sus experimentos únicamente 14 de los 76 atributos disponibles, luego de una intensa investigación se descubrieron algunos otros atributos de suma relevancia para la solución de este problema. Estos estaban siendo ignorados por quienes realizaron las investigaciones para dichos papers.

A continuación se detallan aquellos que luego de la investigación realizada sobre el dominio del problema desprenden más relevancia para la solución del problema.

Attributo	Relevancia	Media	Min	Max	Faltantes
age	La edad es un factor de riesgo, dado que a medida que la persona envejece aumenta las probabilidades de padecer alguna enfermedad relacionada al corazón (Además que en la mayoría de la bibliografía utilizada se remarcaba como uno de los mayores factores de riesgo).	53.481	28	77	0
sex	El riesgo de padecer una enfermedad al corazón es más alto en el hombre que en la mujer. Aunque para esta última, luego de la menopausia aumentan sus probabilidades de padecer alguna enfermedad de este tipo.	0.791	0	1	0
cp	El tipo de dolor en el pecho es un factor de suma relevancia también dado que el tipo de dolor en el pecho podría ser un indicador del tipo de enfermedad del corazón que se padece. Las cuales pueden variar en síntomas y severidad.	3.254	1	4	0
trestbps	La presión sanguínea medida cuando el paciente se encuentra en un estado de reposo constituye un factor importante dado que si la misma es alta significa que alguna vena o arteria se encuentra obstruida por lo cual el corazón necesita aumentar la fuerza con la que bombea la sangre para poder llegar a todo el cuerpo. En el caso de que la misma fuera muy baja podría estar indicando de que el corazón no está funcionando correctamente.	132.259	80	200	60
chol	El colesterol del paciente es sumamente relevante dado que un colesterol alto podría provocar que alguna arteria se obstruya impidiendo al corazón bombear sangre hacia algunas zonas del cuerpo.	247.808	85	603	202
fbs	El nivel de azúcar en sangre puede indicar si el paciente padece de diabetes, enfermedad que aumenta el riesgo de padecer una enfermedad cardíaca.	0.167	0	1	90
restecg	Con esto se puede detectar si el paciente padece alguna condición cardíaca.	0.603	0	2	2
thalach	Un mayor ritmo cardíaco podría indicar que el paciente padece de taquicardia.	137.299	60	202	55
exang	Puede indicar de que el paciente tuvo un preinfarto, en la mayoría de los casos se debe a esto.	0.391	0	1	55
oldpeak	Cuando este valor está descendido (por debajo de 0) indica una isquemia (muerte de tejido por falta de irrigación), y en caso de estar elevado indica un futuro infarto. El segmento ST nos dice el tiempo entre el final de una contracción de los ventrículos, hasta el comienzo de la repolarización (estado de reposo).	0.870	-2600	6200	62
slope	Diferencia los pacientes sin enfermedades coronarias, de los pacientes con enfermedad coronaria de uno, dos, o tres vasos.	1.766	0	3	308
ca	Permite ver la estructura interna del paciente lo cual podría indicar si hubiera alguna obstrucción en algún vaso, vena o arteria.	0.699	0	3	609

thal	Permite indicar si el flujo de sangre dentro del corazón del paciente es normal o no.	5.068	3	7	486
num	Es lo que se busca predecir luego con el uso de nuevos datos, indica si el paciente padece una enfermedad cardíaca o no.	1.129	0	4	0
painloc	El lugar de donde proviene el dolor del pecho ayuda a poder identificar si el paciente padece o podría padecer una enfermedad cardíaca.	0.921	0	1	282
dm	La diabetes es un factor de riesgo, el cual intensifica más su riesgo si el paciente que la padece es mujer. Es una de los causantes de enfermedades cardíacas.	0.958	0	1	804
famhist	Las probabilidades de que un paciente padezca de alguna enfermedad cardíaca aumentan si en su familia alguno la padece o padeció. Es decir si sus predecesores la padecieron o padecen.	0.564	0	1	422
trestbpd	Una anomalía en la presión sanguínea del paciente en estado de reposo podría ser un indicador, por ejemplo, de que este padece de diabetes por ejemplo lo cual conduce a una alto riesgo de padecer una enfermedad cardíaca.	86.623	50	120	60
restwm	Indica si el paciente padece de alguna anomalía en el corazón lo cual, en caso de padecer, podría ser causante de alguna enfermedad cardíaca.	1.033	0	3	869
thalrest	Un ritmo cardíaco anormal en estado de reposo podría ser causa de alguna anomalía cardíaca.	75.488	37	139	56
painexer	El dolor en el pecho durante el ejercicio podría ser provocado por alguna enfermedad cardíaca. Por ende es un factor relevante si se quiere poder predecir si el paciente padece alguna o no.	0.593	0	1	282
relrest	Si el dolor o molestia persiste luego de la actividad física, cuando el paciente ya está descansado, podría indicar la presencia de alguna enfermedad cardíaca.	0.672	0	1	286

Análisis estadístico

Outliers

Para detectar lo Outliers en el dataset se comenzó con los atributos binarios y polinomiales verificando que los valores que tenían en el dataset eran los mencionados en la descripción dada en la página web. Los valores que no se correspondían fueron calificados como outliers y sustituidos por “missings”. Los atributos que contaban con estos errores fueron:

- thal
- ca
- slope

Esto se realizó en el RM con el operador Detect Outlier y Declare Missing Values.

También se detectó casos en los cuales variables continuas tenían valores “0” cuando en realidad deberían tener “missings”, esto se dio en “chol”, en donde un dataset entero tenía valor 0. Este problema también se detectó en “trestbps” y en “trestbps”. Para todos estos casos se reemplazó esos valores por “missings”.

Luego de este análisis previo se comenzó a ver cada una de las variables continuas para evaluar si presentaban valores atípicos o que se escapaban de los valores reales que se podían tener.

Los atributos que se detectaron que contaban con outliers fueron:

- trestbps
- chol
- thalrest

Todos los outliers que se detectaron fueron reemplazados por “missings”.

Reemplazo de Missings

Luego de detectados todos los outliers y transformados a “missings” se continuó con la eliminación de columnas, tuplas o reemplazo de “missings” por valores, como ser la moda o el promedio.

Todas las columnas con más de 200 valores faltantes(de los 900 que tienen los dataset) se quitaron. También se reemplazaron variables con valores “missings”, para esto se tomaron dos criterios: las variables binomiales o polinomiales se reemplazaron por la moda, y las variables continuas se reemplazaron por el promedio. Se optó por no borrar ninguna tupla y reemplazar todos los “missings” por valores.

Esto se realizó con los operadores Select Attribute y Filter Example.

Selección del Método de Aprendizaje

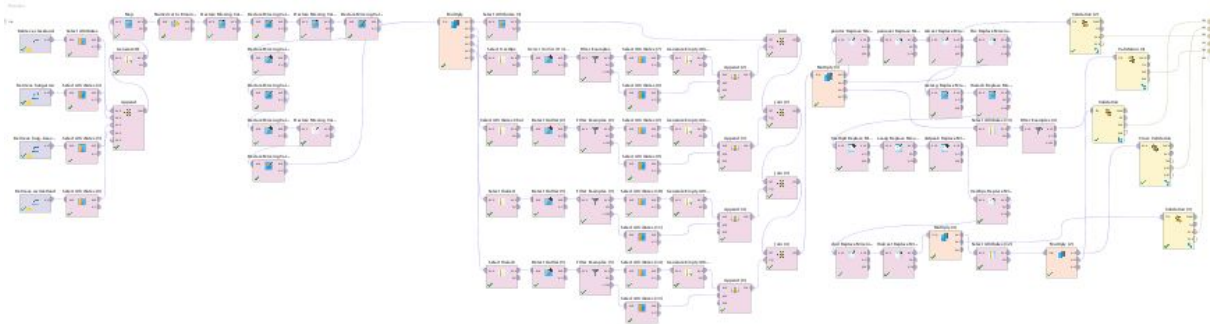
Como método de aprendizaje se optó por utilizar Regresión Logística. El mismo es un método de clasificación cuya respuesta es la probabilidad de pertenencia de la entrada a una de las clases.

La razón por la cual se seleccionó este método y no otro fue simplemente porque se entendía que lo que buscaba el ejercicio era poder predecir si un paciente padece (o padecerá) de una enfermedad cardíaca o no. Por lo tanto estamos ante la presencia de un problema de clasificación binaria (dado que tenemos dos clases, padece o no padece).

Si bien el atributo num, que sería la columna correspondiente a la predicción, puede adoptar cualquier valor entero entre 0 y 4, siendo 0 que el paciente no padece ninguna enfermedad y los restantes 3 cada uno una representación de una distinta enfermedad cardíaca. Se decidió transformar estos valores a binario, es decir 0 no padece, 1 padece.

De esta manera se puede hacer foco en lo que realmente se busca predecir, que es si el paciente padece o padecerá una enfermedad cardíaca (o no).

Diagrama del Proceso en RapidMiner



Conjunto de Entrenamiento y Testing

Utilizamos dos operadores de validaciones para realizar el proceso y poder analizar sus resultados. Los operadores utilizados son los vistos en clase Split Validation y Cross Validation. Para el primero utilizamos un 70% de los casos para entrenamiento, un 30% de los casos para testing y la técnica utilizada es Stratified. Esta técnica nos indica que vamos a realizar los cortes para armar los conjuntos de entrenamiento y testing en “distinto lugar”, es decir que los casos que están en un conjunto para una iteración pueden estar en el otro conjunto o no para la siguiente.

Para el segundo operador utilizamos una validación cruzada de 10 iteraciones, lo que quiere decir que vamos a tener 10 subconjuntos con 100 datos cada uno aprox.

Análisis de los Resultados

Para dar un poco de contexto se explicarán los siguientes dos términos, los cuales se utilizarán luego cuando se analicen los resultados:

- **Precisión:** En el contexto significa el valor de la cantidad de los que reconoció, que efectivamente están bien, sobre la cantidad total de observaciones que reconoció (esto incluye las que se predijo el valor correcto y las que no).
- **Recall:** En el contexto significa la cantidad de las observaciones que se reconoció bien sobre la cantidad total de observaciones que contienen el valor de la predicción.

La siguiente tabla contiene los resultados obtenidos a partir de los cinco distintos experimentos que se hicieron sobre el conjunto de datos. Para cada uno se especifica el método de validación utilizado.

#	Valor	Recall	Precisión	Exactitud
1	true	87.16	83.77	83.64
	false	79.34	83.48	Utilizando: Split Validation
2	true	88.69	77.84	79.87 + - 3.44
	false	69.06	83.28	Utilizando: Cross Validation
3	true	85.81	77.91	78.81
	false	70.25	77.91	Utilizando: Split Validation
4	true	90.54	75.28	78.44
	false	63.4	84.62	Utilizando: Split Validation
5	true	91.86	90.79	80
	false	69.70	72.48	Utilizando: Split Validation

Los dos primeros experimentos fueron los que realizamos para poder analizar la predicción del proceso realizado, el resto son para comparar y verificar los resultados obtenidos. Para el primer experimento utilizamos un split validation de la forma 70-30, es decir 70% de los datos para entrenar y 30% para testear. En este experimento analizamos los missings y los outliers modificando estos como se mencionó anteriormente y obtuvimos la mejor exactitud, 83.64%. Otros puntos que debemos ver son los resultados de precisión y recall, dado que de todos los casos que predijo estuvo bastante acertado un 83.77%, es decir que la precisión es buena, 8 de cada 10 fueron acertados; mientras que se obtuvo un 87.16% de predicción de todos los casos que debió predecir, es decir de 10 que debió predecir, predijo 8. Esto para el conjunto de true, mientras que para el conjunto de false también se obtuvieron resultados acertados siendo estos 83.48% en la precisión y un 79.34% para el recall.

El segundo experimento que utilizamos fue todo igual al anterior, es decir que tomamos todo el detalle del estudio previo del caso, reemplazando missings y quitando outliers pero utilizamos el operador cross validation para la predicción. Los resultados de este experimento fueron un poco peores al experimento anterior, donde obtuvimos un 79.87 +- 3.44 de exactitud.

El resto de los experimentos que tenemos en la tabla son para comparar y verificar si el análisis que realizamos nos da una mejor predicción. Como obtuvimos mejor resultado con el split validation consideramos seguir con este criterio de validación.

El tercer experimento no reemplazamos los missings de los atributos y los resultados fueron peores comparados con el primer experimento donde obtuvimos un 78.81% de exactitud.

El cuarto experimento es igual al primer experimento que realizamos solo que al final no quitamos las columnas que consideramos que tienen mayor número de missings. Haciendo este simple cambio bajamos a un 78.44% la exactitud.

El quinto experimento no reemplazamos los missings de los atributos ,sino que quitamos las columnas que consideramos que tienen mayor número de missings y filtramos para que quite las tuplas que tienen missings. Para este experimento obtuvimos un 80% de exactitud.

Por último seleccionamos solo los 14 atributos que utilizan en todos los papers y los pasamos por nuestro proceso donde obtuvimos los siguientes resultados:

#	Valor	Recall	Precisión	Exactitud
1	true	91.89	79.53	82.53 Utilizando: Split Validation
	false	71.07	87.76	
2	true	89.09	77.92	80.09 + - 2.98 Utilizando: Cross Validation
	false	69.06	83.78	
3	true	79.05	77.91	77.32 Utilizando: Split Validation
	false	75.21	79.59	
4	true	89.86	73.89	76.95 Utilizando: Split Validation
	false	61.6	83.15	
5	true	83.91	74.49	79.26 Utilizando: Split Validation
	false	75.25	84.44	

De esta manera podemos ver que agregando los atributos que consideramos relevantes obtenemos mejores resultados que los obtenidos en papers que hacían referencia a estos datasets y únicamente utilizaban 14 de los 76 atributos.

Referencias

<https://www.nhlbi.nih.gov/health/health-topics/topics/smo>
<http://www.bloodpressureuk.org/BloodPressureandyou/Thebasics/Whatishigh>
<http://www.umm.edu/health/medical/reports/articles/coronary-artery-disease>
<https://www.nhlbi.nih.gov/health/health-topics/topics/hd>
<https://www.nhlbi.nih.gov/health/health-topics/topics/hd/atrisk>
<https://www.hsph.harvard.edu/nutritionsource/disease-prevention/cardiovascular-disease/cv-d-risk-factors/>
http://heartuk.org.uk/files/uploads/documents/huk_fs_mfsl_riskfactorsforchd.pdf
<https://medlineplus.gov/spanish/angina.html>
<http://archive.ics.uci.edu/ml/datasets/heart+Disease>
<https://www.health.harvard.edu/blog/resting-heart-rate-can-reflect-current-future-health-2016-06179806>
<http://www.sabelotodo.org/anatomia/corazon.html>
<https://www.health.harvard.edu/heart-health/chest-pain-a-heart-attack-or-something-else>
<https://estructuradelcorazon.wordpress.com/corazon/miocardio/>
<http://lowngroup.org/how-to-tell-if-your-chest-pain-is-heart-related/>
http://www.heart.org/HEARTORG/Conditions/More/MyHeartandStrokeNews/Family-History-and-Heart-Disease-Stroke_UCM_442849_Article.jsp#mainContent
http://www.heart.org/HEARTORG/Conditions/HeartAttack/SymptomsDiagnosisofHeartAttack/Cardiac-Catheterization_UCM_451486_Article.jsp#mainContent