



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

MATERIA: ANALÍTICA DE DATOS – BIG DATA

NOMBRE DE ESTUDIANTE: Danny Sebastián Díaz Padilla

Laboratorio de:

ANALÍTICA DE DATOS – BIG DATA

Práctica No.: 8

Tema: Análisis del template ‘**Predictive Maintenance**’ de RapidMiner

Objetivos:

- Utilizar un template de RapidMiner.
- Comprender el proceso del template seleccionado de RapidMiner, para entender el problema a detalle.
- Definir los operadores que se utilizan en el proceso del template seleccionado.

Marco teórico:

Template

Es una plantilla que permite guiar, portar, o construir, un diseño o esquema predefinido.

Mantenimiento Predictivo

Modela fallas de equipos basadas en observaciones de referencias de datos de ejecuciones y fallas de máquinas anteriores. El modelo se aplica a situaciones actuales para anticipar las fallas de la máquina y programar el mantenimiento de manera preventiva.

Operador de RapidMiner: Aggregate

Este operador realiza las funciones de agregación conocidas de SQL. Este operador proporciona muchas funcionalidades en el mismo formato que las funciones de agregación de SQL. Las funciones de agregación SQL y las cláusulas GROUP BY y HAVING se pueden imitar con este operador. [1]

Operador de RapidMiner: Append

Este operador crea un conjunto de ejemplos combinado a partir de dos o más conjuntos de ejemplos compatibles agregando todos los ejemplos en un conjunto combinado. [2]



Operador de RapidMiner: Apply model

Este operador aplica un modelo en un conjunto de ejemplos.

Un operador primero entrena un modelo en un ExampleSet, que a menudo es un algoritmo de aprendizaje. Posteriormente, este modelo se puede aplicar en otro ExampleSet. [3]

Operador de RapidMiner: Generate attributes

El operador Generar atributos construye nuevos atributos a partir de los atributos de la entrada ExampleSet y constantes arbitrarias utilizando expresiones matemáticas. Los nombres de atributo de la entrada ExampleSet pueden usarse como variables en las expresiones matemáticas para nuevos atributos. Durante la aplicación de este operador, estas expresiones se evalúan en cada ejemplo, estas variables se rellenan con los valores de atributo del ejemplo. [4]

Operador de RapidMiner: k-NN

Este operador genera un modelo de vecino k-más cercano, que se utiliza para clasificación o regresión.

El algoritmo k-vecino más cercano se basa en la comparación de un ejemplo desconocido con los ejemplos de entrenamiento k que son los vecinos más cercanos del ejemplo desconocido. [5]

Operador de RapidMiner: Log

Este operador almacena información en la tabla de registro. Esta información puede ser casi cualquier cosa, incluidos los valores de los parámetros de los operadores, el recuento de operadores, el tiempo de ejecución, etc. [6]

Operador de RapidMiner: Log to Data

Este operador transforma los datos generados por el operador Log en un conjunto de ejemplos que luego pueden ser utilizados por otros operadores del proceso. [7]

Operador de RapidMiner: Multiply

El operador toma el objeto RapidMiner del puerto de entrada y entrega copias del mismo a los puertos de salida. Cada puerto conectado crea una copia independiente. Por lo tanto, cambiar una copia no tiene efecto en otras copias. [8]



Operador de RapidMiner: Normalize

La normalización se usa para escalar valores para que quepan en un rango específico. Ajustar el rango de valores es muy importante cuando se trata de atributos de diferentes unidades y escalas. Por ejemplo, cuando se usa la distancia euclidiana, todos los atributos deben tener la misma escala para una comparación justa. La normalización es útil para comparar atributos que varían en tamaño. [9]

Operador de RapidMiner: Performance

Este operador se utiliza para la evaluación del desempeño. Ofrece una lista de valores de criterios de rendimiento. Estos criterios de rendimiento se determinan automáticamente para adaptarse al tipo de tarea de aprendizaje. [10]

Operador de RapidMiner: Retrieve

El operador de recuperación carga un objeto RapidMiner en el proceso. Este objeto suele ser un conjunto de ejemplos, pero también puede ser una colección o un modelo. La recuperación de datos de esta manera también proporciona los metadatos del objeto RapidMiner. [11]

Operador de RapidMiner: Reorder Attributes

Este operador permite reordenar los atributos regulares de un conjunto de ejemplos. El reordenamiento se puede realizar alfabéticamente, por especificación del usuario (incluidas las expresiones regulares) o con un conjunto de ejemplos de referencia. [12]

Operador de RapidMiner: Select Attributes

El operador proporciona diferentes tipos de filtros para facilitar la selección de atributos. Las posibilidades son, por ejemplo: Selección directa de atributos. Selección mediante una expresión regular o seleccionando solo Atributos sin valores faltantes. Consulte el *tipo de filtro de atributo de* parámetro para obtener una descripción detallada de los diferentes tipos de filtro. [13]

Operador de RapidMiner: Sort

Este operador clasifica el conjunto de ejemplos proporcionado en el puerto de entrada. El conjunto de datos completo se ordena según un único atributo. Este atributo se especifica utilizando el parámetro de *nombre de atributo*. La clasificación se realiza en dirección creciente o decreciente dependiendo de la configuración del parámetro de *dirección de clasificación*. [14]



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Operador de RapidMiner: Weight by correlation

Este operador calcula la relevancia de los atributos calculando el valor de correlación para cada atributo de la entrada ExampleSet con respecto al atributo de etiqueta. Este esquema de ponderación se basa en la correlación y devuelve el valor absoluto o cuadrado de la correlación como ponderación de atributo. [15]

Operador de RapidMiner: weight to data

Este operador toma los pesos de los atributos como entrada y entrega un conjunto de ejemplos con nombres de atributos y los pesos correspondientes como salida. [16]



Desarrollo de la práctica:

El template seleccionado es el de: Predictive Maintenance



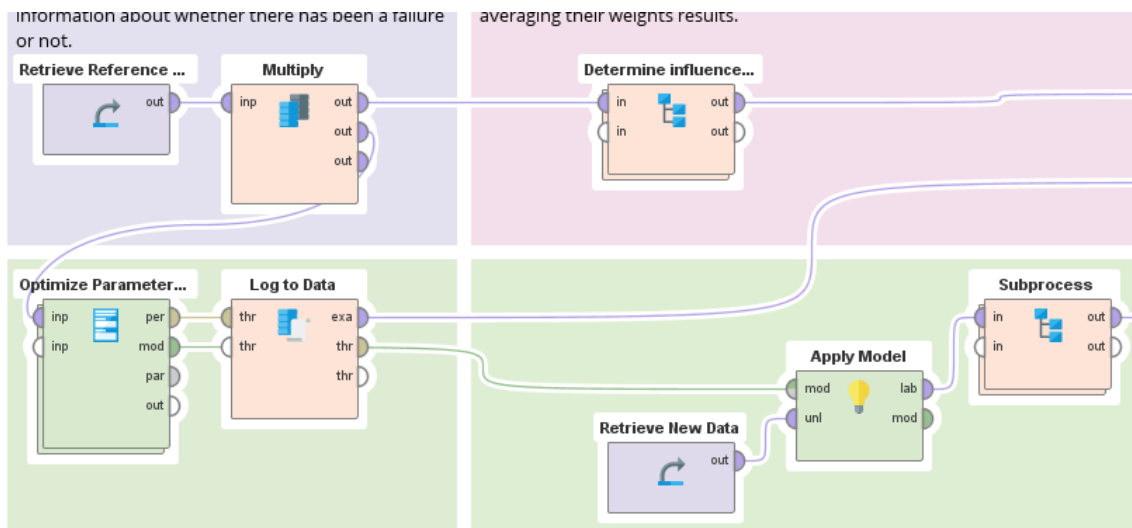
Predictive Maintenance

Model equipment failures to schedule maintenance pre-emptively

A. Proceso

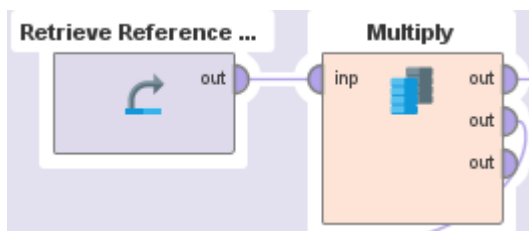
Mantenimiento Predictivo

El esquema del proceso completo es el siguiente:



1. Primer paso:

Se carga los datos de las ejecuciones anteriores de la máquina, etiquetadas con información sobre si ha habido una falla o no.





ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Mediante el operador '**Retrieve**' se carga la base de datos con la información de las máquinas.



Retrieve Reference Data.output (output)

Meta data: Data Table

- Source: //Samples/Templates/Predictive Maintenance/Reference Data

Number of examples = 136

27 attributes:

Note: Some of the nominal values in this set were discarded due to performance reasons. You can change this behaviour in the preferences (rapidminer.

general.md_nominal_values_limit).

Generated by: [Retrieve Reference Data.output](#) ← [Retrieve Reference Data.output](#) ← [Retrieve Reference Data.output](#)

Role	Name	Type	Range	Missings	Comment
d	Machine_ID	nominal	=[M_0001, M_000...	= 0	
label	Failure	binominal	=[no, yes]	= 0	
	Sensor_1	real	=[0.211 - 11.243]	= 0	
	Sensor_2	real	=[0.545 - 32.498]	= 0	
	Sensor_3	real	=[0.497 - 31.421]	= 0	
	Sensor_4	real	=[0.586 - 32.322]	= 0	
	Sensor_5	real	=[0.600 - 31.946]	= 0	
	Sensor_6	real	=[0.589 - 26.361]	= 0	
	Sensor_7	real	=[0.346 - 19.681]	= 0	
	Sensor_8	real	=[0.136 - 15.341]	= 0	
	Sensor_9	real	=[0.349 - 12.207]	= 0	
	Sensor_10	real	=[0.131 - 9.666]	= 0	
	Sensor_11	real	=[0.158 - 5.713]	= 0	

Se toma esta entrada y se agrega al operador '**Multiply**' el cual crea una copia de la información enviándola al paso 2 y al paso 3.

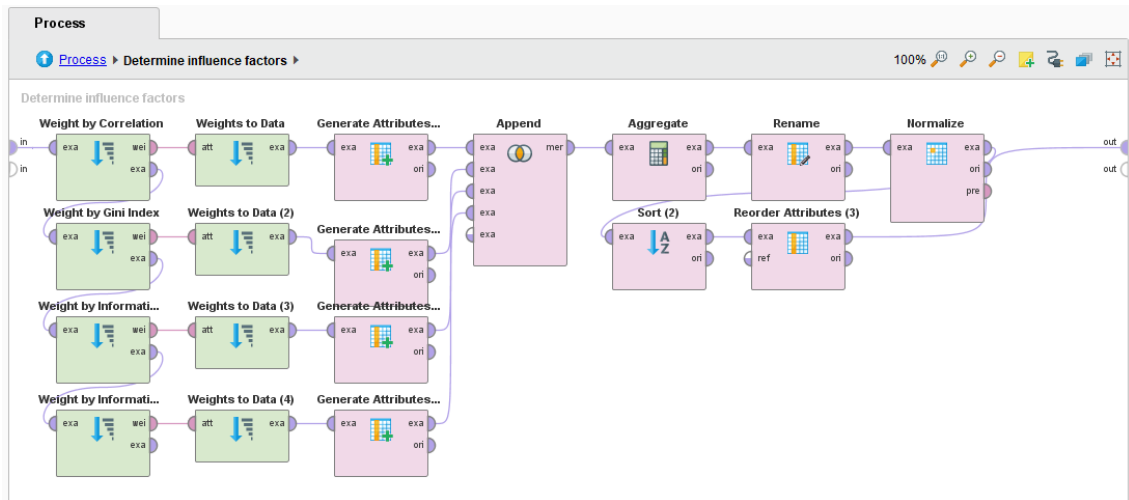
2. Segundo paso:

Se determina los factores de influencia utilizando varios algoritmos de ponderación de atributos y promediando sus resultados de ponderaciones.

Usando la entrada del primer paso, se realiza un subproceso que se crea un peso (relevancia) por cada atributo de acuerdo a: la correlación índice Gini, ganancia de información e índice de ganancia de información.



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN



Luego para cada peso se genera un conjunto de datos de ejemplo, el operador Generar atributos construye nuevos atributos a partir de los atributos de la entrada y constantes arbitrarias utilizando expresiones matemáticas.

Luego usando el operador 'Append' se combinan todos los conjuntos de ejemplos.

A este conjunto combinado se le aplica una función de agregación (Aggregate) para obtener el promedio de los pesos.

aggregation attribute	aggregation functions
Weight	average

Y le da un nuevo nombre usando el operador 'Rename':

old name	new name
average(Weight)	Importance



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Y por un segundo canal envía la información original que recibió (sin realizar la agregación) para ordenarla en forma descendente por **importancia**.

Parameters

Sort (2) (Sort)

attribute name: Importance

sorting direction: decreasing

Y finalmente redistribuir sus atributos.

Parameters

Reorder Attributes (3) (Reorder Attributes)

sort mode: user specified

attribute ordering: Define order...

use regular expressions: ☐

handle unmatched: remove

sort direction: ascending

Attribute Ordering

Importance

La salida de este paso son estos dos conjuntos de datos ya explicados.

Determine influence factors.out 1 (out 1)

Meta data: Data Table

Number of examples = 0

2 attributes:

Generated by: [Determine influence factors.out 1](#)

Role	Name	Type	Range	Missings	Comment
	Attribute	nominal	unknown	= 0	
	Importance	real	= [0 - 1]	= 0	

Además, su resultado es final, es decir, ya no pasa por más procesos.

Dando como salida:

Views: Design Results Turbo Prep A

ExampleSet (Select Attributes)

ExampleSet (Reorder Attributes (3))

Result History

Log

Open in: Turbo Prep Auto Model

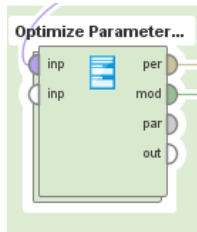
Row No.	Attribute	Importance
1	Sensor_7	1
2	Sensor_6	0.918
3	Sensor_8	0.877
4	Sensor_5	0.807
5	Sensor_2	0.790



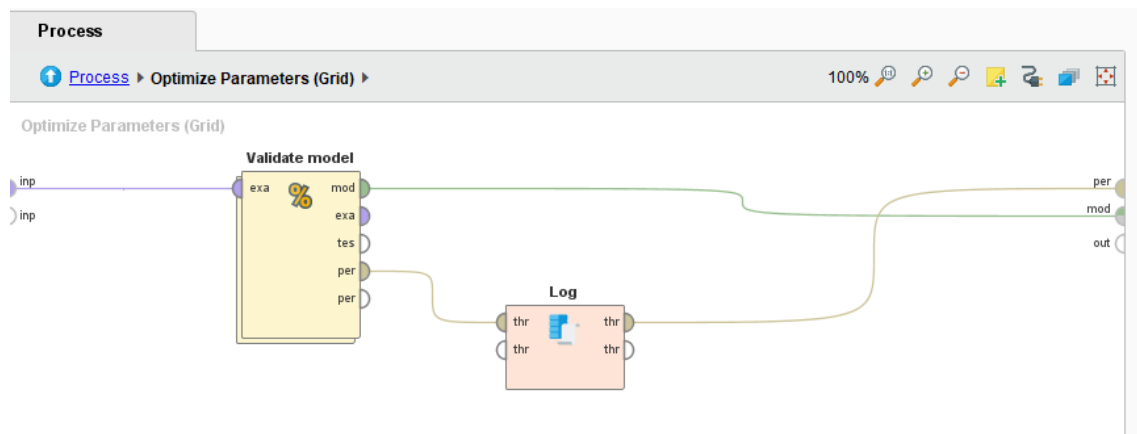
ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

3. Tercer paso:

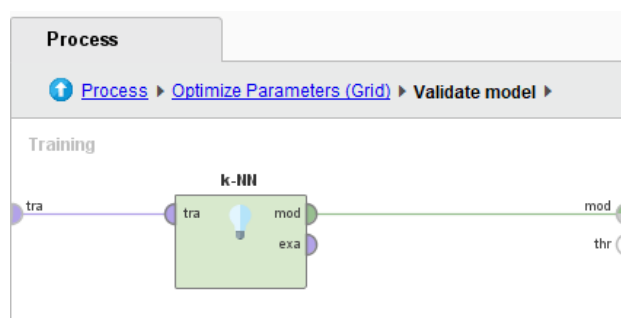
Se entrena un modelo k-NN: se optimiza para k (el número de situaciones de referencia a tener en cuenta para la predicción) para producir una precisión de predicción de falla máxima.



Usando la salida del primer paso como entrada, se crea un proceso interno que validará el modelo.



Dentro del proceso interno del subprocesso se usa el operador 'k-NN' para realizar el entrenamiento del modelo, operador que genera un modelo de k vecinos más cercano, que se utiliza para clasificación o regresión. Se basa básicamente en la comparación de un ejemplo desconocido con los ejemplos de entrenamiento k que son los vecinos más cercanos del ejemplo desconocido. Los cuales corresponden a una parte de la entrada del primer paso.



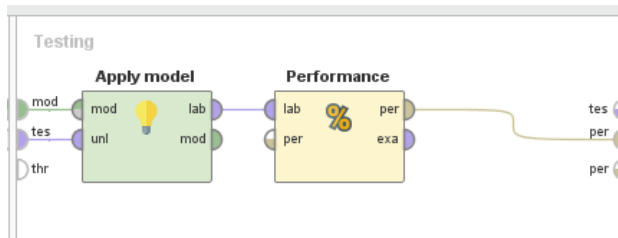
La salida de esta parte es el modelo ya entrenado.



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

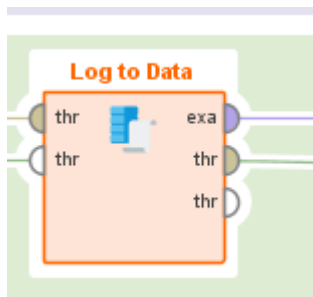
Para la parte del testeo se usa el operador **'Apply model'** para aplicar el modelo sobre un subconjunto de testeo previamente obtenido de la entrada del primer paso.

Y luego evalúa el desempeño del modelo usando el operador **'Performance'** la única salida de esta parte es las medidas de desempeño



Finalmente se utiliza el operador **'Log'** para recuperar las medidas de desempeño.

La salida de este subprocesso consiste en un modelo entrenado y en las mediciones de desempeño.



Finalmente usando el operador **'Log to Data'** se genera un conjunto de ejemplos usando la información del **'Log'** previamente recuperado.

Row No.	iteration	k-NN.k	accuracy
1	1	1	0.619
2	13	13	0.631
3	26	26	0.684
4	27	27	0.684
5	14	14	0.670
6	2	2	0.587
7	28	28	0.691
8	15	15	0.682
9	3	3	0.671
10	16	16	0.647
11	4	4	0.633
12	29	29	0.704
13	30	30	0.707
14	31	31	0.706



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

La salida de este paso consiste en la data generada desde el log y la salida del subproceso anteriormente detallado.

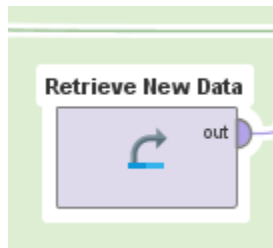
Además, el conjunto de ejemplos es un resultado final.

4. Cuarto paso:

Se carga nuevos datos y se aplica el modelo de falla de la máquina a las ejecuciones actuales de la máquina para predecir posibles fallas.

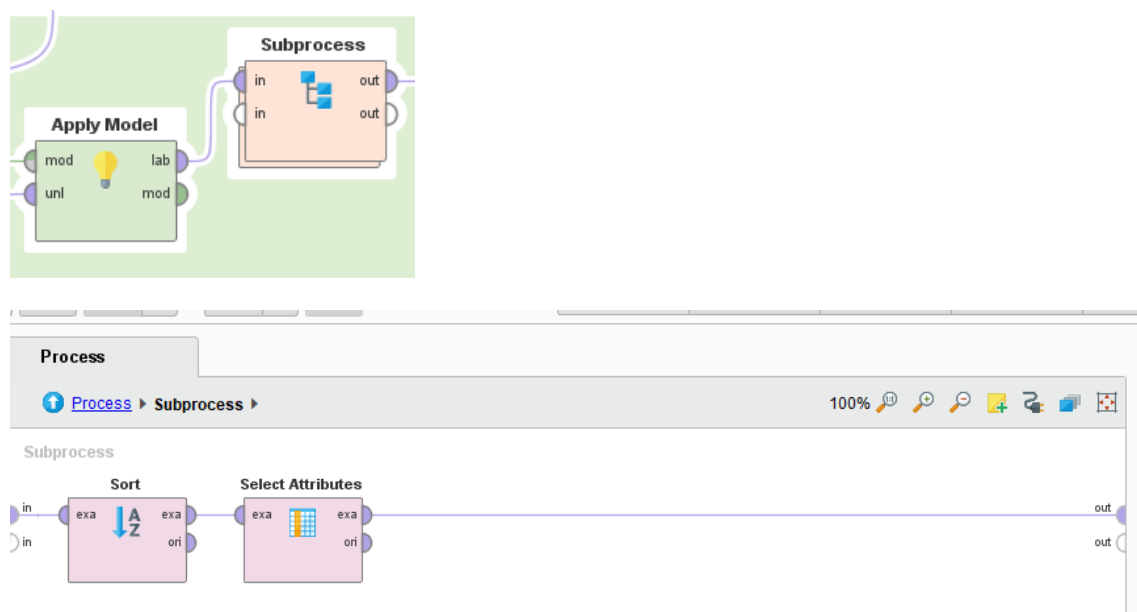
Primero se carga de nuevo los datos con el operador **'Retrieve'**

Role	Name	Type	Range	Missings
id	Machine_ID	nominal	$\neq [M_0001, M_0002] = 0$	
	Sensor_1	real	$\neq [0.211 - 11.243] = 0$	
	Sensor_2	real	$\neq [0.545 - 32.498] = 0$	
	Sensor_3	real	$\neq [0.497 - 31.421] = 0$	
	Sensor_4	real	$\neq [0.586 - 32.322] = 0$	
	Sensor_5	real	$\neq [0.600 - 31.946] = 0$	
	Sensor_6	real	$\neq [0.589 - 26.361] = 0$	
	Sensor_7	real	$\neq [0.346 - 19.681] = 0$	
	Sensor_8	real	$\neq [0.136 - 15.341] = 0$	
	Sensor_9	real	$\neq [0.349 - 12.207] = 0$	
	Sensor_10	real	$\neq [0.131 - 9.666] = 0$	
	Sensor_11	real	$\neq [0.158 - 5.713] = 0$	
	Sensor_12	real	$\neq [0.052 - 3.323] = 0$	
	Sensor_13	real	$\neq [0.014 - 1.968] = 0$	
	Sensor_14	real	$\neq [0.039 - 1.404] = 0$	
	Sensor_15	real	$\neq [0.007 - 3.139] = 0$	



Se aplica el modelo usando **'Apply model'** al conjunto de datos generados desde el Log utilizando el modelo generado en el tercer paso.

Y luego por medio de un subproceso se trata la data generada.





ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Ordenando a los datos con **'Sort'** por confiabilidad de forma decreciente

Parameters

Sort

attribute name: confidence(yes)

sorting direction: decreasing

Y seleccionando todos los atributos utilizando **'Select Attributes'**:

Parameters

Select Attributes

attribute filter type: all

☒ invert selection

☐ include special attributes

Finalmente, la salida de las predicciones de fallas

views: Design Result

ExampleSet (Select Attributes)

Result History Log

Open in Turbo Prep Auto Model

Row No.	Machine_ID	prediction(F...	confidence(...	confidence(...
1	M_0223	yes	0.100	0.900
2	M_0239	yes	0.100	0.900
3	M_0176	yes	0.133	0.867
4	M_0194	yes	0.133	0.867
5	M_0221	yes	0.133	0.867
6	M_0225	yes	0.133	0.867
7	M_0240	yes	0.133	0.867
8	M_0250	yes	0.133	0.867
9	M_0166	yes	0.167	0.833
10	M_0229	yes	0.167	0.833
11	M_0271	yes	0.167	0.833
12	M_0137	yes	0.200	0.800
13	M_0140	yes	0.200	0.800
14	M_0151	yes	0.200	0.800



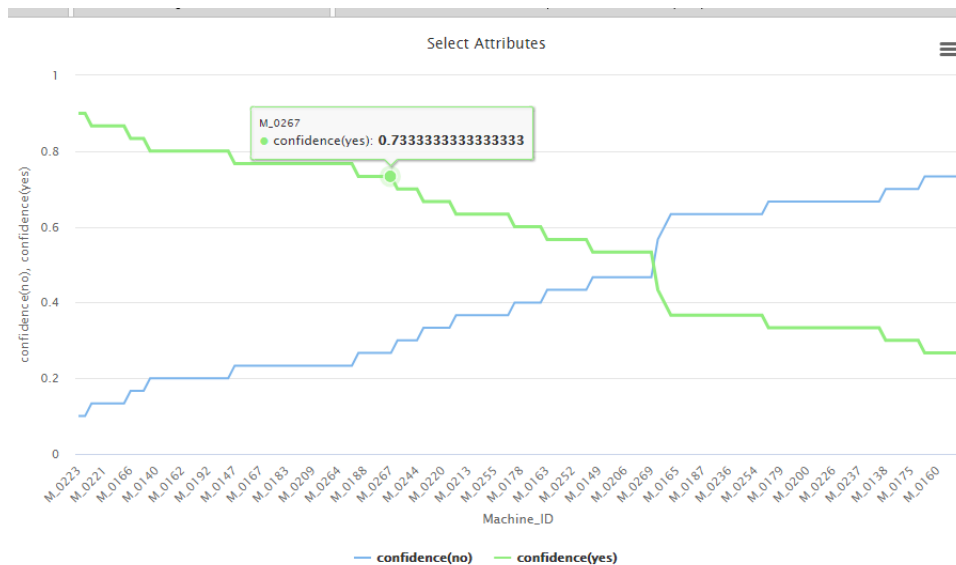
ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

B. Resultados estadísticos

	Name	Type	Missing	Statistics	Filter (4 / 4 attributes): <input type="text" value="Search for Attributes"/>
Data	Machine_ID	Nominal	0	Least M_0136 (0) Most M_0137 (1)	Values M_0137 (1), M_0138 (1), ...[270 more]
Statistics	Prediction prediction(Failure)	Binominal	0	Least no (47) Most yes (89)	Values yes (89), no (4) Details...
Visualizations	Confidence_no confidence(no)	Real	0	Min 0.100 Max 0.733 Average 0.421 Deviation 0.199	Open visualizations
Annotations	Confidence_yes confidence(yes)	Real	0	Min 0.267 Max 0.900 Average 0.579 Deviation 0.199	Open visualizations

C. Visualizaciones

En este caso se determina que hay un cambio brusco cuando la confiabilidad está alrededor de 0.5.

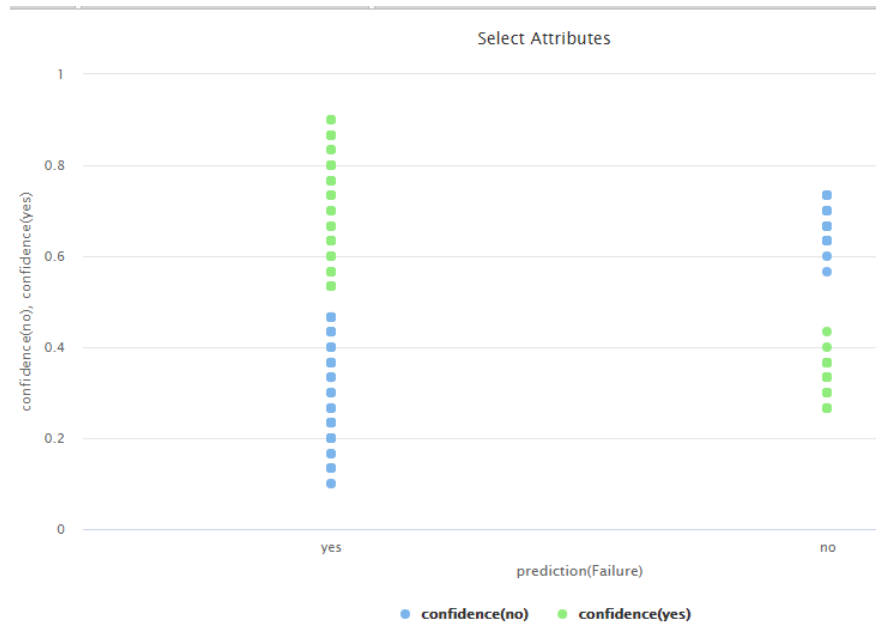




ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Análisis de resultados:

El conjunto de datos poseía más información sobre máquinas que efectivamente han fallado.



Por lo que puede influir en la predicción el no tener los datos de forma balanceada. Creando un sesgo cognitivo en los pesos generados por correlación.

En general, el proceso es capaz de generar una predicción bastante adecuada además de generar conjuntos de datos nuevos a partir de los metadatos obtenidos de los pesos por correlación de todos los datos de las máquinas que fallaron y no fallaron.

Conclusiones y recomendaciones:

- Se utilizó un el template 'Predictive Maintenance' de RapidMiner, para analizar la relación entre mediciones de sensores sobre una máquina y sus fallos.
 - Se logró entender y explicar a detalle todo el proceso del template seleccionado de RapidMiner.
 - Se han definido puntualmente con la documentación de RapidMiner todos los operadores que se utilizan en el proceso del template seleccionado.
 - Se realizó un análisis de pesos por correlación y con estos pesos se generó nueva data.
 - Se usó un entrenamiento K-NN, para determinar la correlación de los k vecinos más cercanos.
-
- Se recomienda utilizar una base de datos con categorías balanceadas en lo posible para no crear sesgos en la predicción.
 - Es recomendable agrupar los procesos en subprocesos (Como se vio en este template) para mejor entendimiento del grupo de trabajo.



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Bibliografía

- [1]"Aggregate - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/blending/table/grouping/aggregate.html>. [Accessed: 27- Nov- 2019].
- [2]"Append - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/blending/table/joins/append.html>. [Accessed: 27- Nov- 2019].
- [3]"Apply Model - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/scoring/apply_model.html. [Accessed: 27- Nov- 2019].
- [4]"Generate Attributes - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/blending/attributes/generation/generate_attributes.html. [Accessed: 27- Nov- 2019].
- [5]"k-NN - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/lazy/k_nn.html. [Accessed: 27- Nov- 2019].
- [6]"Log - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/utility/logging/log.html>. [Accessed: 27- Nov- 2019].
- [7]"Log to Data - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/utility/logging/log_to_data.html. [Accessed: 27- Nov- 2019].
- [8]"Multiply - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/utility/multiply.html>. [Accessed: 27- Nov- 2019].
- [9]"Normalize - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/cleansing/normalization/normalize.html>. [Accessed: 27- Nov- 2019].
- [10]"Performance - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/validation/performance/performance.html>. [Accessed: 27- Nov- 2019].
- [11]"Retrieve - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/data_access/retrieve.html. [Accessed: 27- Nov- 2019].
- [12]"Reorder Attributes - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/blending/attributes/order_attributes.html. [Accessed: 27- Nov- 2019].
- [13]"Select Attributes - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/8.0/studio/operators/blending/attributes/selection/select_attributes.html. [Accessed: 27- Nov- 2019].
- [14]"Sort - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/blending/examples/sort/sort.html>. [Accessed: 27- Nov- 2019].
- [15]"Weight by correlation - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_correlation.html. [Accessed: 27- Nov- 2019].
- [16]"Weight to data - RapidMiner Documentation", *Docs.rapidminer.com*. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weights_to_data.html. [Accessed: 27- Nov- 2019].