



Modelamiento y Análisis Predictivo 3ª parte: Regresión Multivariable con Treenet®

En su afán de aportar contenido orientativo a los profesionales que se dedican a la gestión de los procesos con propósitos de mejoramiento de la productividad, Quality, entrega esta 3ª y última publicación sobre el módulo de analítica predictiva que está disponible en el software Minitab en su versión 21. Sin embargo, los algoritmos que se utilizan están disponibles en las bibliotecas de herramientas de Machine Learning para cualquiera que los necesite utilizar, por ejemplo vía programación con el lenguaje Python.



Darío Alarcón Hunter
Director

Es muy probable que nuestros lectores se estén preguntando:

¿y esto en qué podría ayudarme para la operación de mis procesos productivos o transaccionales?

Esperemos que podamos dar respuesta a esas inquietudes, desde el contenido multimedial de estas tres publicaciones en serie, sobre la temática de la Analítica Predictiva y su facilidad para desarrollar el modelo de un

procesos muy complejo o de muchas variables, con el apoyo de los datos históricos

Vamos avanzando **¿Cuál es el principal desafío que enfrenta cualquier operador de un proceso?** Conocerlo a tal profundidad, que, ante un evento inusual, tenga la capacidad de anticiparse a los resultados a lograr o dicho de otro modo, lo conozca tan bien que pueda modelar y predecir su comportamiento con relativa certeza.

La idea parece atractiva; pero hasta hace poco, era tremendamente laborioso de lograr, dado que sólo se tenía acceso a las herramientas tradicionales, como las regresiones simples, que eran muy restringidas o herramientas más complejas aún, como las regresiones multivariantes no lineales, que exigen mucha creatividad. Siendo así y tras muchas horas de trabajo, si se llegaba a lograr un modelo con un coeficiente de determinación o R^2 superior a 75%, podía considerarse muy satisfecho, ya que eso significaba que el 75% de la variabilidad de la variable objetivo de salida, se explicaba por el comportamiento de los predictores analizados.

Felizmente en la actualidad, se cuenta con algoritmos muy eficaces empaquetados en un software, que otorgan mucha facilidad a los ingenieros; para configurar las bases de un modelo y el software se encarga de entregar un modelo, también empaquetado, que viabiliza esa respuesta de predicción que se necesita **¿cómo se comportará el proceso si tengo este nuevo conjunto de variables?** con estas nuevas herramientas, tendrá su respuesta con relativa facilidad y valores de R^2 superiores a 90%, claro que eso será muy dependiente de la cantidad de registro de datos que utilice.

Si no ha tenido la oportunidad de conocer esta nueva opción de analítica predictiva y dado que estos son temas de reciente divulgación, le sugerimos acceda previamente a <https://www.quality.cl/minitab-21-y-la-regresion-cart-de-analitica-predictiva/> en la que se tratan gradualmente: la regresión lineal múltiple, la regresión no lineal multivariable clásica y se finaliza con la

CART® es ahora estándar en MINITAB 21

presentación de la herramienta **CART**[®] (Classification and Regression Trees). En la eventualidad que ya esté familiarizado con el uso y aplicación de **CART**[®], puede salvar tiempo e ir directo a su temática de interés.

Si tiene inconvenientes en acceder al link de **CART**[®] ya indicado, el video también está disponible en nuestro canal de Youtube “Grupo Quality”, seleccione <https://youtu.be/yEU90LdAgWg> y en 32 minutos (21 minutos si lo configura a la velocidad de 1,5 entendible para hispano parlantes) lo guiará por los temas ya mencionados, creando árboles de decisión para una respuesta continua con muchas variables predictoras categóricas y continuas.

Introducción a CART

Paso 1: Crecimiento de un gran árbol
Se hace en forma automática

Todas las variables se consideran en cada división del árbol

Cada división se hace usando una sola variable a un valor específico ($x_1 = 55.3$) o a un conjunto de valores ($x_1 <= 78.4$)

Las divisiones se escogen de modo de minimizar el error del modelo

El árbol sigue creciendo hasta cumplir un criterio especificado por el analista o hasta que no pueda crecer más.

Paso 1: Podar el árbol
En función de lo que configura el analista

Se utiliza la opción "Selección un Árbol Alternativo" o utilizando muestras de ensayo o la validación cruzada para definir los criterios para podar los sub-árboles

© 2011 Dirección Regional de Quality College

Regresión Simple (Revisión) - 17

Como ya fue presentada en la segunda sección de esta serie, si le interesa como opera un algoritmo como Random Forest, le invitamos a ver el contenido de nuestro video <https://youtu.be/Pnlx6PwODqE>, que en 25 minutos (20 minutos si lo configura a la velocidad 1,25 claramente entendible para hispano parlantes) le permitirá visualizar como lograr un modelo con el algoritmo **Random Forest**[®]; que no es un prerrequisito, para el entendimiento del modelamiento con **TreeNet**[®], como si lo es el de **Cart**[®].

Random Forest de Molienda Primaria Presentación del Análisis

Importancia relativa de variables

Variable	Importancia relativa (%)
L2 INH	70.0
UM SAS 02 (N)	17.2
Amo SAS 02 (N)	10.2
Publ SAS 02 (N)	1.0
CE SAS 02 (N)	1.4
Clas de SAS 02	0.3
CE BA SAS 02 (N)	0.9
VAL SAS 02 (N)	0.2
Clas de SAS 02	0.1
Pos SAS 02 (N)	0.1

La importancia de las variables mide la mejora del modelo cuando se realizan divisiones en un predictor. La importancia relativa se define como un % de mejora con respecto al predictor estandar.

© 2011 Dirección Regional de Quality College

Machine Learning - 19

Finalmente, les invitamos a visitar el tercero y último de nuestros videos sobre modelos de analítica predictiva <https://youtu.be/FWxBzZ-4kdQ> en el que se presenta el análisis con el algoritmo **TreeNet**[®], que por su grado de elaboración, permite entregar

Regresión Treenet de Pilas de Lixiviación Presentación del Análisis, para 1.242 árboles

Parcela de superficie de Rend. Cierre ajustado

Tratado de contorno de Rend. Cierre ajustado

En el gráfico en 3D se observa una depresión que sugiere observar en paralelo con el gráfico de contorno de la derecha.

La gráfica de contorno nos permite ver con mayor claridad la inconformidad de usar áreas más altas superiores a 10 mil (m²), midiendo en % de la SAS 02 por sobre el 1% tal como se muestran los gráficos de 1 predictor.

© 2011 Dirección Regional de Quality College

Regresión Treenet - 39

análisis adicionales como las gráficas de dependencia parcial de un predictor respecto de otros predictores de importancia para el modelo, en 2 y en 3 dimensiones, lo que resulta clave para entender las interacciones entre los distintos predictores. Los invitamos a suscribirse a nuestro canal, para recibir oportunamente nuestras notificaciones.

Tal como lo hicimos para los casos prácticos en la sección de **Random Forest**[®]; mostramos el mismo caso de las probetas de concreto y también dos de los modelos de procesos mineros que escogimos para mostrar las bondades del enfoque de la analítica predictiva:

Modelo	$S=\sqrt{MSE}$	R^2
Regresión Random Forest	3436	97,55
Regresión Treenet	2267	98,9

datos, el cuadro adjunto, muestra algunos de esos resultados comparativos.

- Desempeño de la Producción de molino SAG, ante la presencia de 11 predictores, en la que se utilizan 1003 registros de

Modelo	$S=\sqrt{MSE}$	R^2
Regresión Random Forest	4,8	81,1 %
Regresión Treenet	2,4	96,2 %

predictores, fundamentalmente de dimensionamiento físico de las pilas.

- Análisis del Rendimiento de 523 Pilas de Lixiviación. Se necesitaba predecir su comportamiento frente a 10

Esperamos que con este contenido pueda formarse una idea de las herramientas que están disponibles para que un profesional, medianamente entrenado en el uso de estas herramientas de analítica predictiva, esté en condiciones de proponer en horas, modelos muy eficaces de los procesos productivos o transaccionales de cualquier organización

La gran ventaja, es que no se tienen restricciones sobre la cantidad de predictores a utilizar; no requiere la identificación previa de las interacciones entre predictores; no es necesario eliminar los datos faltantes o atípicos, etc. Estos nuevos algoritmos empaquetados en el Módulo de Analítica Predictiva, disponibles en Minitab 21, le podrán facilitar tremendamente su trabajo.

Estamos atentos para proponerle agendas de cursos que se puedan adecuar a sus necesidades específicas o para apoyarlo en el análisis de datos; si su necesidad fuera más inmediata <http://www.quality.cl>