



Analítica Predictiva 2: Regresión Multivariable con Random Forest®

Quality, desde 1990 está apoyando a las organizaciones nacionales -con servicios de capacitación y consultoría en los ámbitos de sistemas de gestión y mejoramiento de la productividad de los procesos productivos y de servicios- y una vez más les comparte un artículo, que persigue nuestro permanente afán de acercar el análisis estadístico y la ciencia de los datos a los profesionales que necesitan aplicarlas; para conocer, entender y mejorar los desempeños de los procesos operativos o transaccionales por los cuáles responden, con el propósito de agregar valor a las organizaciones en las que colaboran.



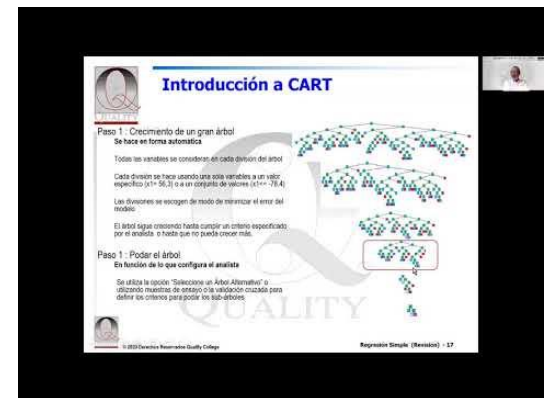
Darío Alarcón Hunter
Director

En la eventualidad que sea la primera vez que conoce de la 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, pasando por la regresión no lineal multivariable clásica y finalizando con la presentación de la herramienta **CART®** (Classification and Regression Trees), que ha sido incluido a partir de MINITAB 19 como parte

CART® es parte de MINITAB 21 como nuevo estándar de herramientas disponibles

integral de las herramientas disponibles entre muchas otras, para el análisis de datos, bajo la Etiqueta “Analítica Predictiva” que incluye “**Clasificación CART**” para procesos con variables de salida discretas (no tratado en este artículo todavía) y también “**Regresión CART**” si es que el proceso tiene una o varias variables de salida continuas.

Otra opción es que acceda en forma directa a nuestro canal en Youtube y seleccione nuestro video <https://youtu.be/yEU9OLdAgWg> que en 32 minutos (21 minutos si lo configura a la velocidad de 1,5 entendible para hispano parlantes) lo va a guiar conceptualmente por los temas ya mencionados, creando árboles de decisión para una respuesta continua con muchas variables predictoras categóricas y continuas, apoyado con ejemplos desarrollados, para proponerle un modelo; que le permitirá predecir el comportamiento de la variable objetivo del proceso, en base a los valores previstos de los predictores a incluir.



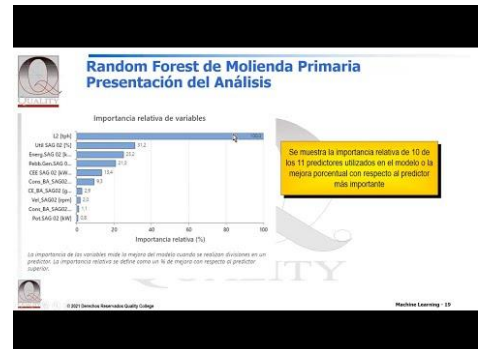
El propósito de este nuevo artículo es avanzar en otro de los algoritmos disponibles en el módulo de analítica predictiva pero

El “Módulo de Analítica Predictiva” adiciona los algoritmos Random Forest® y TreeNet®

que se entrega como un adicional, bajo la forma “Módulo de Analítica Predictiva” que por ahora pone a disposición los algoritmos de Machine Learning **Random Forest®** y también **TreeNet®**, que será tratado con posterioridad, como una 3ª parte de esta temática de Analítica predictiva. Como ya se ha

indicado estos algoritmos están disponibles tanto para variables objetivo de tipo discreta como de tipo continua, pero nosotros trataremos por ahora sólo variables objetivo de tipo continua.

Aprovechando las opciones multimediales, 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 opera el algoritmo **Random Forest**[®], que se basa en las propiedades ya indicadas en nuestro artículo Analítica Predictiva 1 de **CART**[®], que utilizando un algoritmo validado y patentado por MINITAB permite mejorar el R² de los modelos abordados o dicho de otro modo mejorar el desempeño de la predicción **CART**[®].



No es nuestro interés en enfatizar o profundizar en las bondades o particularidades de un algoritmo, por el contrario nos interesa su aplicabilidad en el día a día y es por esa razón que nuestro video, continúa con el desarrollo del análisis de los datos de probetas de

Modelo	$S=\sqrt{MSE}$	R ²
Regresión CART (1 árbol con 175 nodos)	6,45	85,05
Regresión Random Forest (100 árboles y 700 muestras)	5,76	88,6

concreto que ya habíamos trabajado con **CART**[®], pero que ahora se muestra con otro modelamiento predictivo, utilizando **Random Forest**[®]. Lo anterior permite descubrir en forma comparativa, las

potencialidades de este nuevo algoritmo; que sin duda presenta un desempeño superior debido a la menor varianza lograda al trabajar con múltiples árboles de predicción y no sólo 1 como en **CART**[®]. Por supuesto; esto presenta mucha facilidad a cualquier analista que sea competente en la configuración del algoritmo, para evaluar los desempeños de distintos modelos de predicción; **sin la necesidad de un experto en ciencia de datos.**

En forma adicional a los datos de concreto, que se tratan nuevamente en este nuevo video; quisimos ir más allá y mostrar ejemplos de lo apropiado de este tipo de algoritmo a la realidad minera; presentando la predicción de 2 nuevos modelos de **Regresión Multivariantes**, utilizando **Random Forest**[®], en los que se podrá apreciar que lo que se necesita hacer; es configurar sólo los parámetros del algoritmo y Minitab 21, entrega una serie de tablas, gráficos y opciones de predicción -también configurables, que le permitirán al analista mediante una simple observación de los entregables, entender entre otros, cuáles son los predictores de mayor impacto en el modelo investigado.

Los modelos mineros escogidos fueron los siguientes:

- Desempeño de la Producción de un molino SAG, ante la presencia de 11 predictores, en la que se utilizan 1003 registros de datos.
- Análisis del Rendimiento de 523 Pilas de Lixiviación. Se necesita predecir su comportamiento frente a 10 predictores, fundamentalmente de dimensionamiento físico.

Les invitamos a visitar el video que le proponemos y estaremos atentos a satisfacer sus consultas o comentarios que le parezcan pertinentes.

En nuestro tercer artículo de esta serie de analítica predictiva -que se publicará los próximos días en nuestro sitio y también en nuestro canal- abordaremos el mismo tipo de análisis, pero utilizando esta vez, el algoritmo **TreeNet**[®], que por su grado de elaboración, permite entregar 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 predictores. Los invitamos a suscribirse a nuestro canal, para recibir oportunamente esa notificación.

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