



Smart BI

Inteligencia Estratégica de Negocios



- › **Conceptos Básicos de Machine Learning**
- › **Tipos de Machine Learning**
- › **Machine Learning en la Industria Petrolera**
- › **Como Implementar Machine Learning en la Industria Petrolera**
- › **Ejemplos de Machine Learning en la Industria Petrolera**



Machine Learning

¿Qué es?

Es el campo de estudio que da a las computadoras la **capacidad de aprender** sin haber sido programadas explícitamente.

¿Para qué sirve?

Para crear **modelos** que **aprendan** constantemente y que ofrezcan valores **futuros** estimados cada vez más precisos.

¿Por qué es importante?

Porque brinda a las organizaciones la capacidad de **resolver** problemas complejos y poder hacer **predicciones** sobre valores totalmente nuevos.

El Machine Learning (ML) está revolucionando nuestro mundo de formas inimaginables.

En la era digital en la que vivimos estamos rodeados de datos en cantidades masivas. Estos datos contienen información valiosa que puede ayudarnos a tomar decisiones más inteligentes, a resolver problemas complejos y a descubrir patrones ocultos.





En lugar de programar reglas específicas, **entrenamos a los algoritmos** para que **encuentren patrones** en los datos y **realicen predicciones** o tomen decisiones basadas en esos patrones.

Este **enfoque revolucionario** ha demostrado ser **extremadamente útil** en una amplia gama de aplicaciones. **El ML está transformando la forma en que interactuamos con el mundo.**



Aprendizaje Supervisado

En este enfoque, se proporciona al Algoritmo un conjunto de **datos de entrenamiento** que incluye las **entradas** y las **salidas esperadas**. El algoritmo aprende a partir de estos datos y luego se le puede presentar un conjunto de datos nuevo para **realizar predicciones o clasificaciones**.

Aprendizaje No Supervisado

A diferencia del Aprendizaje Supervisado, en el Aprendizaje No Supervisado **no se proporcionan salidas esperadas** al algoritmo. En cambio, el algoritmo **busca patrones y estructuras ocultas** en los datos por sí mismos. Es útil cuando **no tenemos conocimiento previo** de las clase o categorías en los datos.

Aprendizaje Semi-Supervisado

Infiere una función de datos combinados, donde **algunos están etiquetados** (se conoce la variable objetivo) **y otros no** están etiquetados (no se conoce la variable objetivo).

Aprendizaje por Refuerzo

La computadora aprende a través de **ensayo y error**, mejora continuamente su toma de decisiones basándose en la **retroalimentación** de su entorno .

Cada uno de estos tipos de aprendizaje emplea uno o más modelos estadísticos:

Regresión. Modelo que produce un resultado numérico.

Clasificación. Modelo que produce un resultado –texto, o no numérico–.

Clustering. Modelo que crea grupo de datos con características comunes.

Asociación. Modelo que crea relaciones entre datos que previamente han tenido un patrón.



La cantidad de datos en el sector de Oil & Gas ha aumentado de forma exponencial conjuntamente con el avance de la tecnología de información.

Esto incluye desde el registro de sensores en operaciones de exploración, perforación, producción y sísmica, hasta la tecnología de Logging While Drilling (LWD), que permite registrar los datos de perforación en tiempo real.



El ML traerá la siguiente transformación en la Industria Petrolera, ya que hay una gran variedad de aplicaciones en las que **puede ser implementada exitosamente**, ayudando a mejorar la productividad, la seguridad y las operaciones.

El ML puede permitir a las empresas **optimizar** el diseño, la perforación y la terminación de pozos, **pronosticar con alta precisión** la producción de petróleo de un campo, **anticipar problemas** de funcionamiento, fallos o fugas de ductos que generen algún tipo de riesgo.



1. Definir el Objetivo del Proyecto. Tener una clara comprensión de lo que deseas lograr con el Machine Learning. Definir el problema que deseas resolver y establecer metas claras y medibles.

2. Recopilar y Preparar los Datos. Obtener los datos necesarios para entrenar el modelo de Machine Learning. Asegurar tener suficientes datos de calidad y prepararlos adecuadamente, lo que puede incluir limpiarlos, normalizarlos y dividirlos en conjuntos de entrenamiento, validación y pruebas.

3. Seleccionar y Entrenar el Modelo. Elegir el algoritmo de Machine Learning adecuado para el proyecto y entrenar el modelo utilizando los datos de entrenamiento. Ajustar los parámetros del modelo para obtener el mejor rendimiento posible.

4. Evaluar el Modelo. Evaluar el rendimiento del modelo utilizando los datos de validación. Utilizar métricas de evaluación apropiadas para el problema, como precisión, recall o error cuadrático medio.

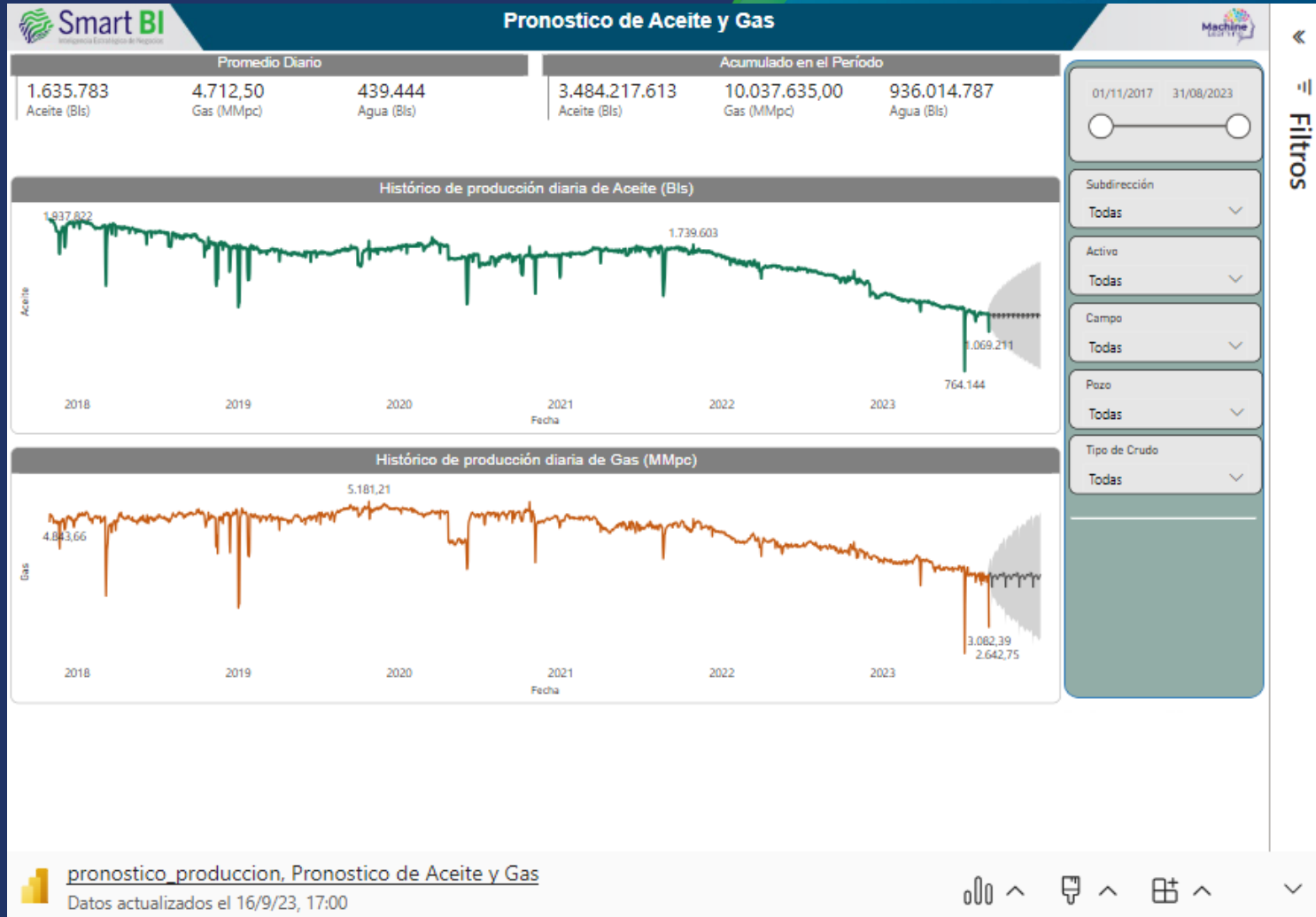
5. Ajustar y Optimizar el Modelo. Si el rendimiento del modelo no es satisfactorio, realizar ajustes y optimizaciones. Probar diferentes algoritmos, ajustar los parámetros o utilizar técnicas de preprocesamiento de datos adicionales.

6. Validar el Modelo. Una vez que se está satisfecho con el rendimiento del modelo, validar su eficiencia utilizando datos de prueba. Esto dará una idea más realista de cómo se comportará en situaciones reales.

7. Implementar y Monitorear el Modelo. Implementar el modelo en producción y monitorear su rendimiento continuamente. Realizar ajustes según sea necesario y mantener el modelo actualizado con nuevos datos.

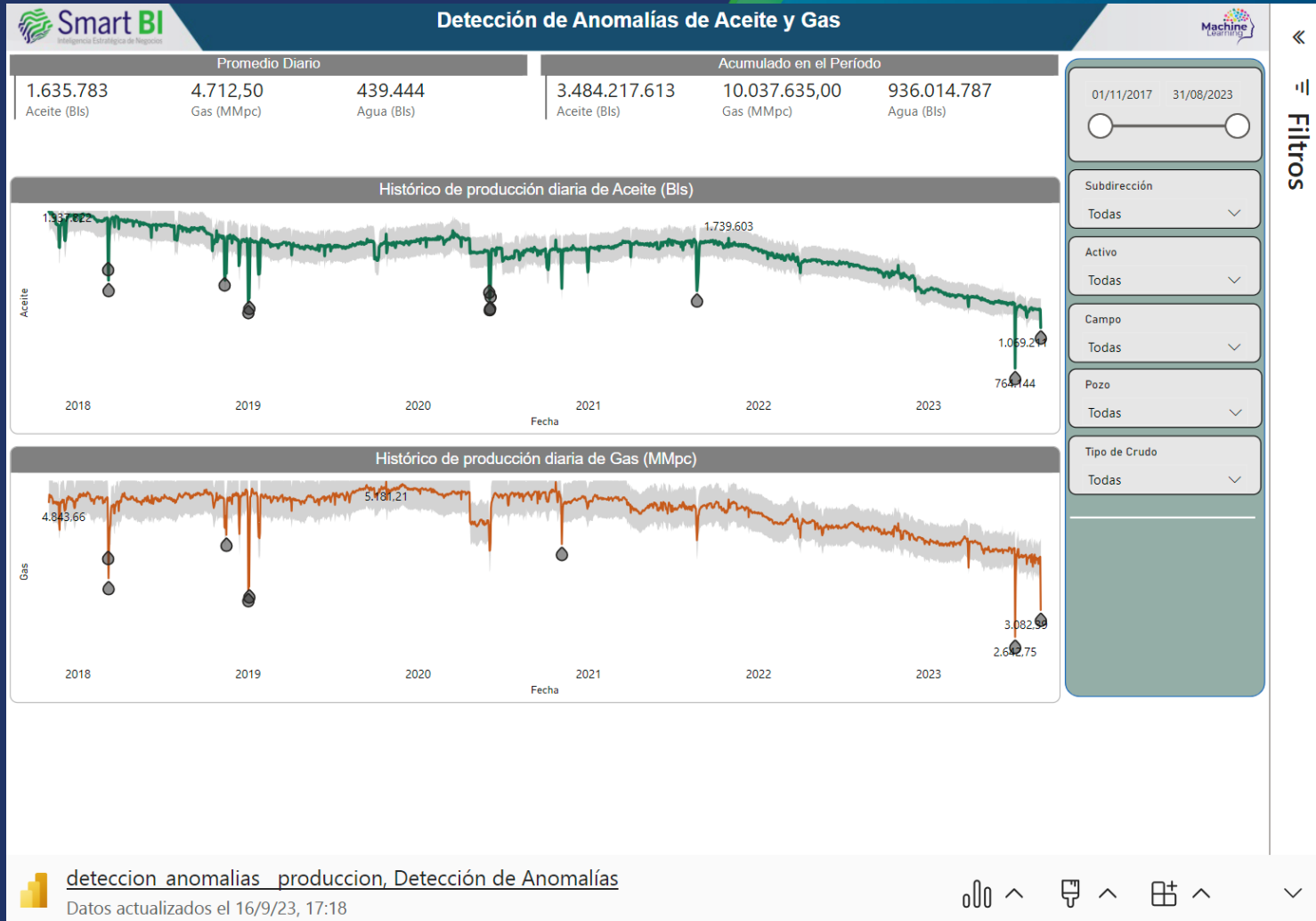
Pronósticos de Producción.

Su propósito es hacer pronósticos de la producción de aceite, gas y agua, basándose en los datos históricos. Power BI, dispone de la característica de forecasting en series de tiempo.



Detección de Anomalías.

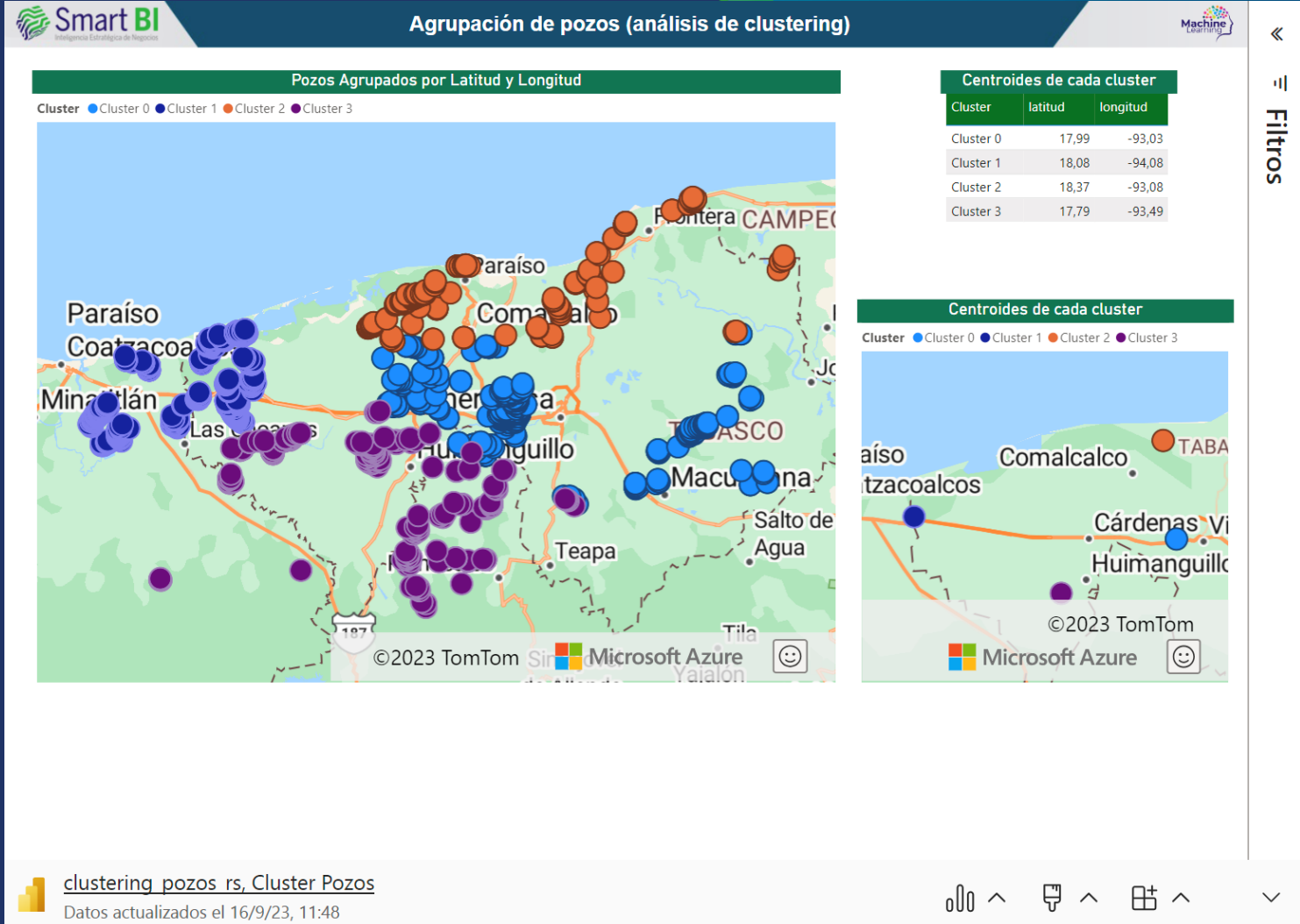
Su propósito es identificar valores fuera de rango o valores atípicos y ver que los está causando. Power BI, dispone de la característica de detección de anomalías en series de tiempo.



Análisis de Clustering – Análisis geoespacial.

Su propósito es agrupar pozos con latitudes y longitudes similares, en el ejemplo se muestra una agrupación de pozos con solo tener su latitud y longitud.

Hacemos uso del modelo kmeans de la librería Pycaret de Python para obtener las agrupaciones, posteriormente se calculan los centroides de cada grupo.



Análisis de Regresión – Tasa de Penetración

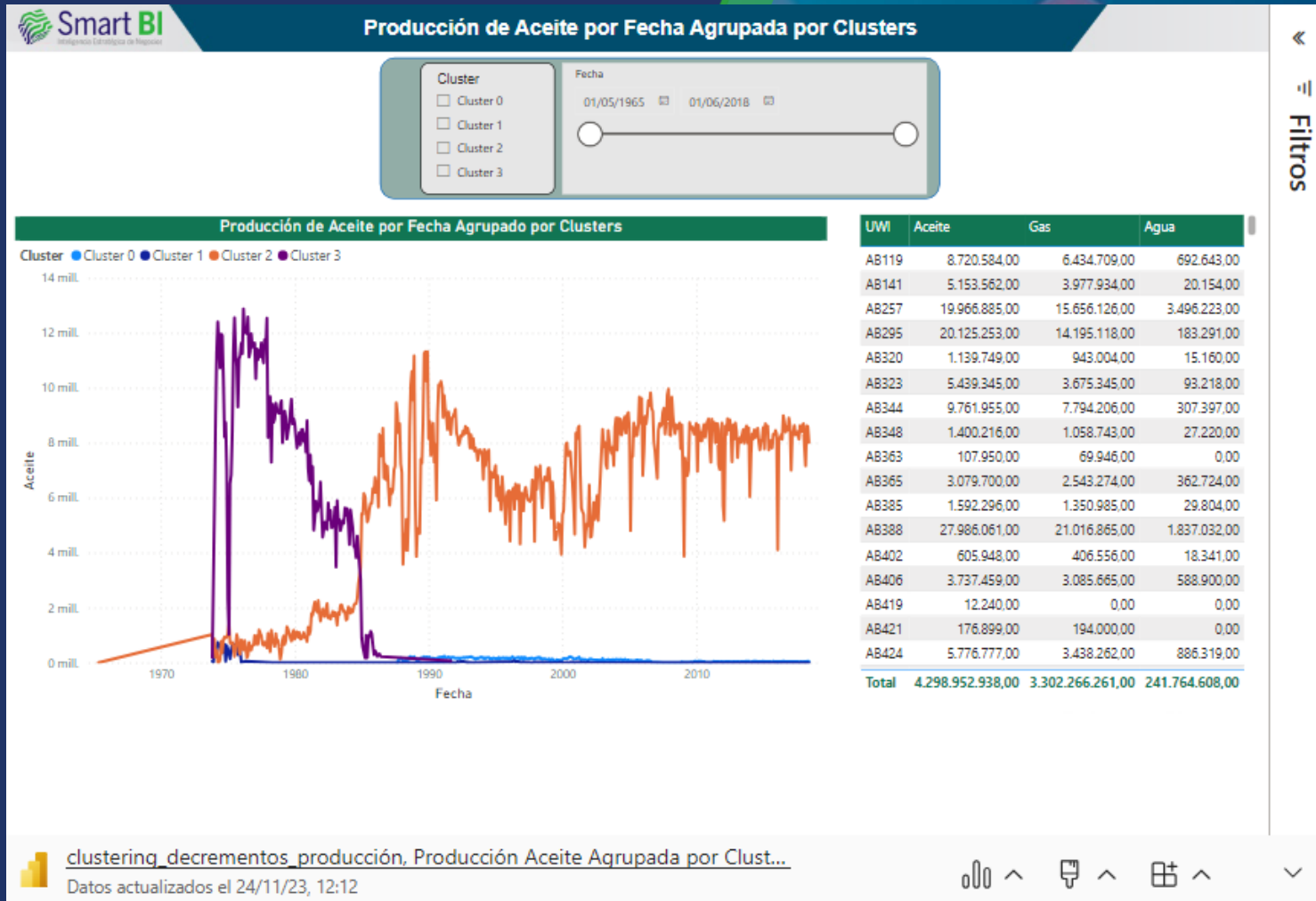
Su propósito es predecir la tasa de penetración de perforación (ROP), que es la velocidad a la cual la sarta o broca perfora la roca subyacente para profundizar al pozo. Se utiliza un conjunto de datos con la siguiente información: presión diferencial, carga de gancho, peso en la broca, entre otros.



Análisis de Clustering – Decrementos de Producción.

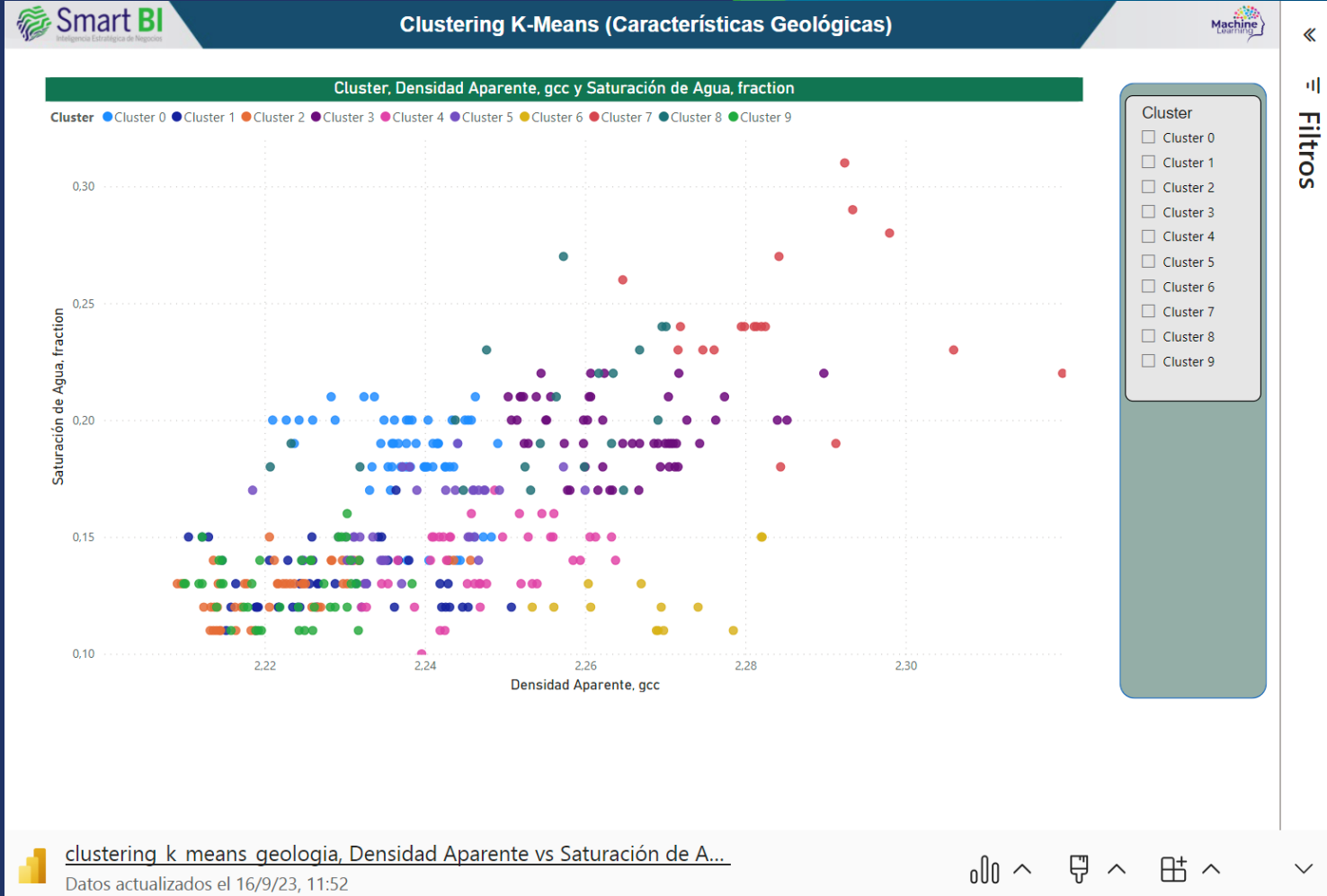
Su propósito es agrupar pozos con producción histórica de aceite gas y agua.

Hacemos uso del modelo kmeans de la librería Pycaret de Python para obtener las agrupaciones. Una vez que el modelo es ejecutado se pueden ver los pozos agrupados por sus producciones e identificar pozos con una alta probabilidad de disminución de producción en el futuro.



Análisis de Clustering – Características Geológicas de pozos.

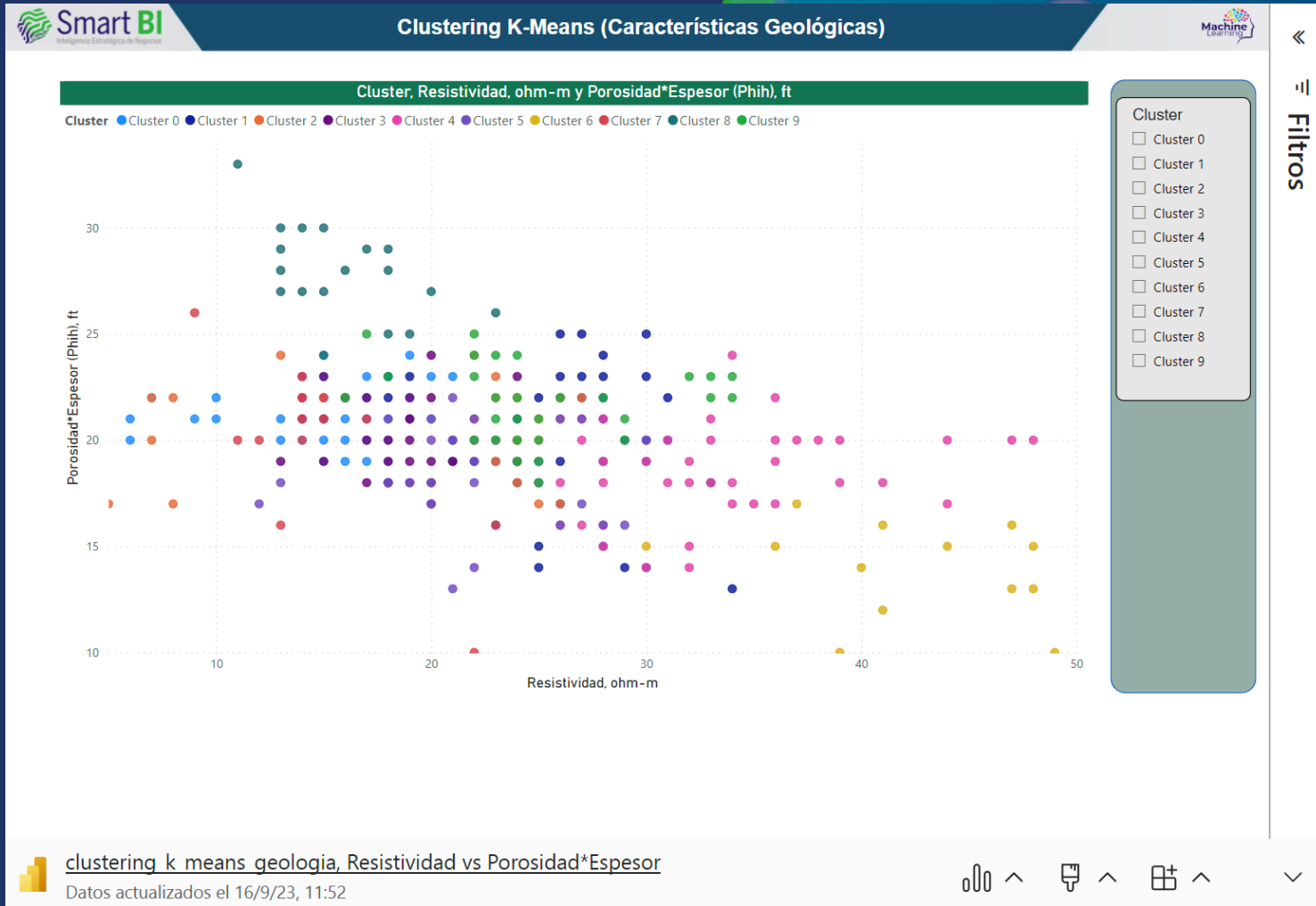
Su propósito es agrupar pozos con características geológicas similares, en el ejemplo se muestra un diagrama de dispersión entre la densidad y la saturación de agua de varios pozos y se observa cómo se agrupan pozos con características similares. Hacemos uso del modelo kmeans de la librería Pycaret de Python para obtener las agrupaciones.



Análisis de Clustering – Características Geológicas de pozos.

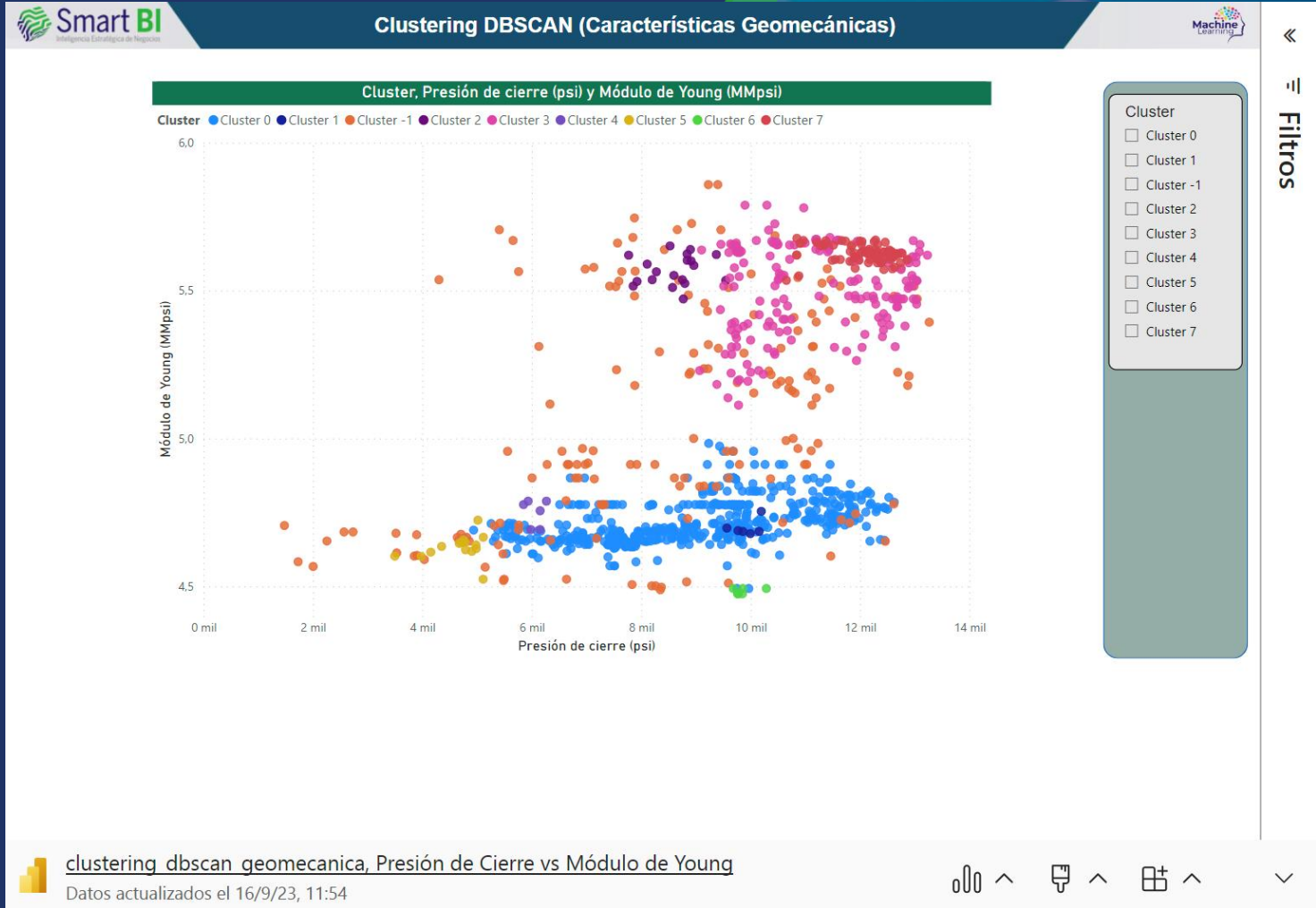
Su propósito es agrupar pozos con características geológicas similares, en el ejemplo se muestra un diagrama de dispersión entre la resistividad y la porosidad*espesor de varios pozos y se observa cómo se agrupan pozos con características similares.

Hacemos uso del modelo kmeans de la librería Pycaret de Python para obtener las agrupaciones.



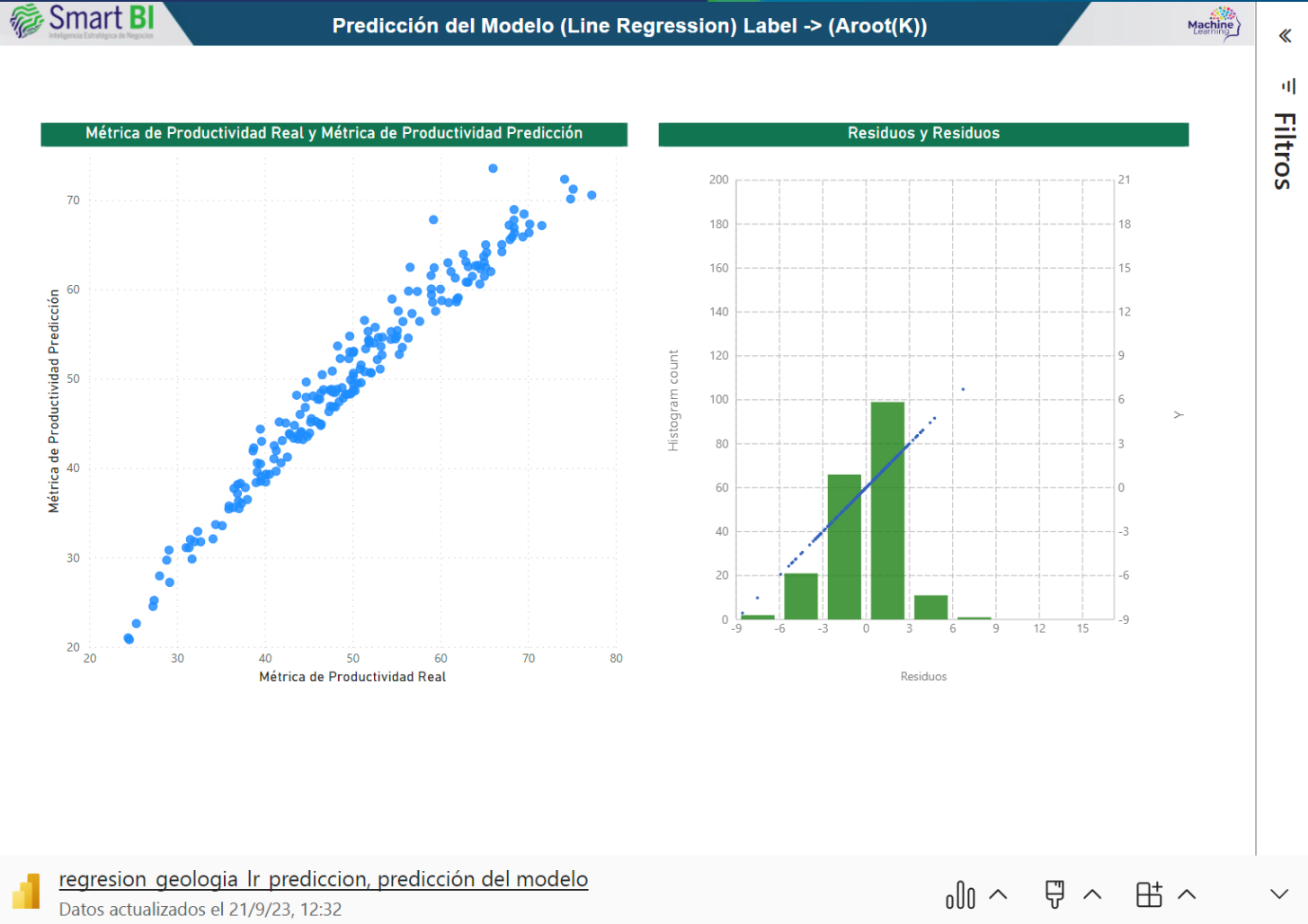
Análisis de Clustering – Características Geomecánicas de pozos.

Su propósito es agrupar pozos con características geomecánicas similares, en el ejemplo se muestra un diagrama de dispersión entre la presión de cierre y el módulo de Young, observe los valores atípicos del cluster -1. Hacemos uso del modelo dbscan de la librería Pycaret de Python para obtener las agrupaciones.



Análisis de Regresión Multilineal – Rendimiento de Productividad.

Su propósito es predecir los valores de rendimiento de la productividad a partir de un conjunto de datos con información geológica, en el ejemplo se muestra un diagrama de dispersión entre el rendimiento de la productividad real vs al rendimiento de la productividad predicha por el modelo. Hacemos uso del modelo LR de la librería Pycaret de Python para obtener las predicciones.



Análisis de Regresión Multilineal – Rendimiento de Productividad.

Su propósito es predecir los valores de rendimiento de la productividad a partir de un conjunto de datos con información geológica, en el ejemplo se presenta una tabla con las columnas del conjunto de datos y se muestra la productividad real y la productividad predicha por el modelo. Hacemos uso del modelo LR de la librería Pycaret de Python para obtener las predicciones.

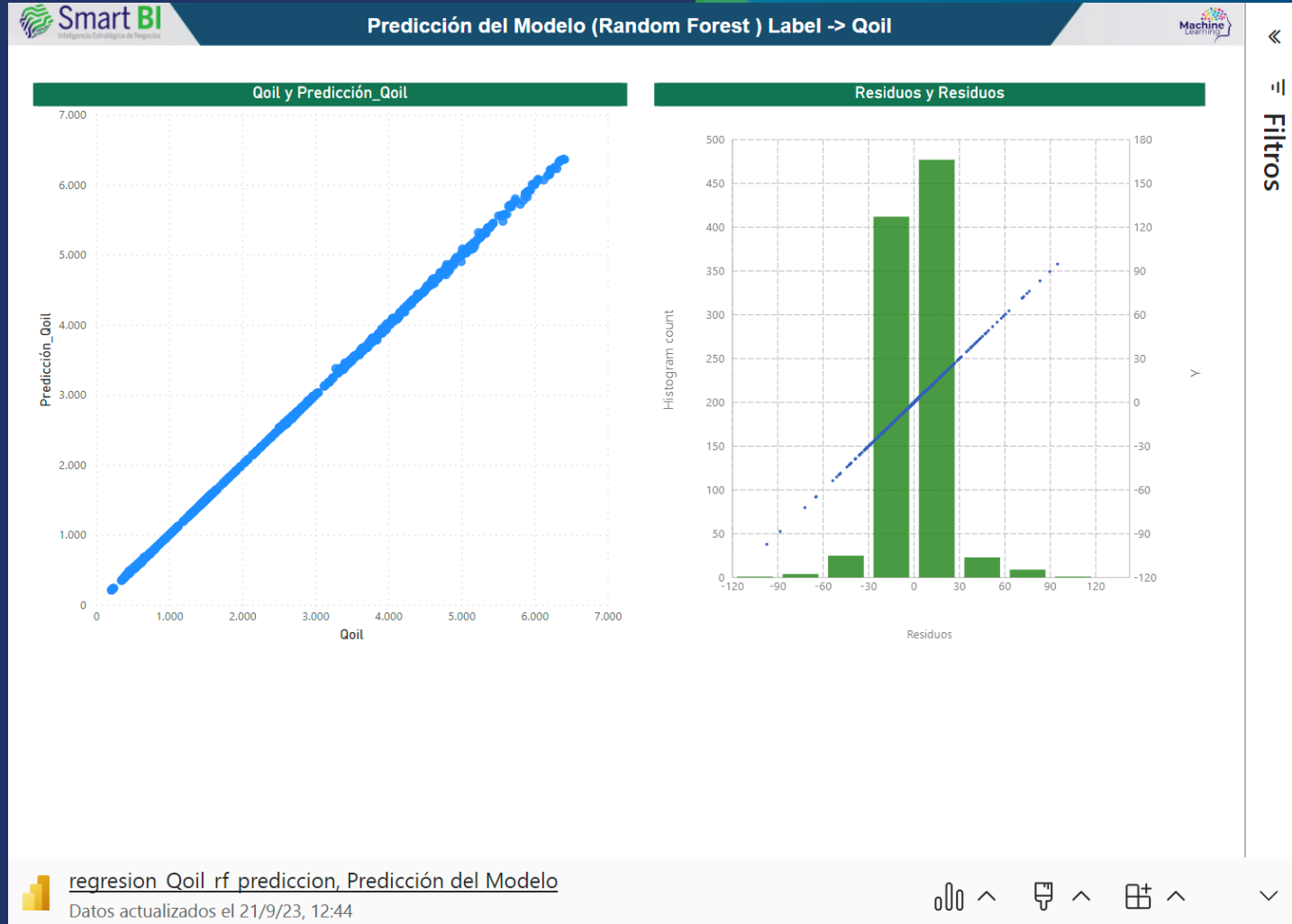
Smart BI **Predicción del Modelo (Productividad Real vs Productividad Predicción)** Machine Learning

Acoustic impedance (kg/m2s*10^6)	Brittleness Ratio	Matrix Perm (nd)	Porosity (%)	TOC (%)	Vitrinite Reflectance (%)	Metrica de Productividad Real	Metrica de Productividad Predicción	Exactitud del modelo
3,77	69,64	272	7,94	2,28	1,72	41,06	41,05	99,99
2,01	72,16	349	11,63	6,16	1,49	60,02	60,02	99,99
2,05	40,96	422	12,52	7,00	1,30	54,39	54,40	99,97
3,75	48,29	401	9,60	3,36	1,74	44,14	44,10	99,90
3,66	75,53	172	7,52	1,24	1,51	37,79	37,83	99,89
2,48	41,58	216	10,47	4,28	1,34	43,30	43,35	99,89
3,95	69,64	349	8,78	1,88	1,62	43,69	43,63	99,87
2,52	41,74	349	10,70	5,12	1,29	45,23	45,15	99,83
3,48	35,54	256	7,76	3,32	1,57	31,05	31,11	99,81
2,66	43,01	410	11,28	5,92	1,47	49,77	49,88	99,78
3,39	42,59	841	13,69	5,36	1,56	62,93	63,10	99,73
2,99	71,18	466	12,43	6,68	1,88	65,18	64,99	99,71
3,59	29,71	346	8,30	0,32	1,10	28,02	27,93	99,69
3,21	71,16	394	9,92	3,64	1,53	50,11	50,31	99,61
3,48	45,64	505	11,30	3,88	1,64	50,86	51,08	99,57
4,19	44,56	382	11,16	5,44	1,94	48,80	49,01	99,57
3,89	58,58	253	8,73	3,60	2,03	44,11	43,91	99,54
4,69	74,27	378	7,10	0,15	1,58	35,93	35,77	99,53
3,77	68,32	322	8,79	3,16	1,82	45,32	45,53	99,52
3,42	69,56	505	10,63	4,60	1,73	55,11	55,39	99,50
3,62	67,19	371	10,72	5,40	1,94	55,07	54,78	99,47

regresion geologia lr prediccion, Productividad Real y Productividad Pr...
Datos actualizados el 21/9/23, 12:32

Análisis de Regresión Random Forest – Predicción de Qoil.

Su propósito es predecir los valores de Qoil en un pozo de flujo natural, con mas de 2000 pruebas de flujo disponible, en el ejemplo se muestra un diagrama de dispersión entre el Qoil real vs al Qoil predicha por el modelo. Hacemos uso del modelo RF de la librería Pycaret de Python para obtener las predicciones.



Análisis de Regresión Random Forest – Predicción de Qoil.

Su propósito es predecir los valores de Qoil en un pozo de flujo natural, con mas de 2000 pruebas de flujo disponible, en el ejemplo se presenta una tabla con las columnas del conjunto de datos y se muestra el Qoil real y el Qoil predicha por el modelo. Hacemos uso del modelo RF de la librería Pycaret de Python para obtener las predicciones.

Smart BI Inteligencia Estratégica de Negocios

Predicción del Modelo (Qoil vs Predicción Qoil)

Machine Learning

Filtros

Presión de la Cabeza del Pozo	Temperatura de la Cabeza del Pozo	Presión del Orificio Inferior	Temperatura del Separador	Diámetro Interno Estrangulador	Presión del Separador	Qoil	Predicción_Qoil	Exactitud del modelo
3.946,11	80,08	5.453,92	60,27	0,25	100,00	605,60	604,23	99,77
3.940,92	79,91	5.455,15	60,27	0,25	200,00	600,60	599,45	99,81
3.900,17	79,54	5.457,84	60,26	0,25	300,00	589,40	592,31	99,51
3.843,18	80,57	5.447,57	60,23	0,25	100,00	648,30	641,66	98,96
3.836,74	80,41	5.448,76	60,23	0,25	200,00	643,30	638,20	99,20
3.739,40	79,73	5.453,81	60,22	0,25	300,00	621,80	619,38	99,61
3.725,73	81,08	5.440,85	60,19	0,25	100,00	695,80	691,03	99,31
3.709,16	83,23	5.429,01	60,35	0,25	100,00	547,60	561,93	97,45
3.704,22	80,85	5.442,58	60,19	0,25	200,00	688,30	677,89	98,46
3.689,65	82,98	5.430,83	60,34	0,25	200,00	541,70	554,98	97,61
3.667,09	78,11	5.468,20	60,24	0,25	400,00	546,50	553,86	98,67
3.603,12	92,83	5.355,71	61,43	0,32	100,00	1.007,40	1.007,40	100,00
3.603,12	92,83	5.355,71	61,43	0,32	200,00	1.007,40	1.007,40	100,00
3.603,12	92,83	5.355,71	61,43	0,32	300,00	1.007,40	1.007,40	100,00
3.603,12	92,83	5.355,71	61,43	0,32	400,00	1.007,40	1.007,40	100,00
3.603,12	92,83	5.355,71	61,43	0,32	500,00	1.007,40	1.007,40	100,00
3.603,12	92,83	5.355,71	61,43	0,32	600,00	1.007,40	1.007,40	100,00
3.601,08	83,82	5.422,60	60,31	0,25	100,00	580,80	582,21	99,76
3.590,92	81,60	5.433,86	60,15	0,25	100,00	748,50	752,79	99,43
3.565,79	93,10	5.348,17	61,56	0,32	100,00	1.063,90	1.063,90	100,00
3.565,79	93,10	5.348,17	61,56	0,32	200,00	1.063,90	1.063,90	100,00

regresion Qoil rf prediccion, Qoil y Qoil Predicción

Datos actualizados el 21/9/23, 12:44

Análisis de Regresión Lineal - Predicción de Tiempo de Viaje de la Onda de Compresión.

Su propósito es predecir los tiempos de viaje de una onda de compresión, estos tiempos por lo regular se obtienen de los registros sísmicos, en este ejemplo se entrena a un modelo de aprendizaje automático utilizando regresión lineal para hacer dicha predicción.

Smart BI **Predicción del Modelo (LR) Label -> Tiempo de Viaje de la Onda de Compresión** Machine Learning

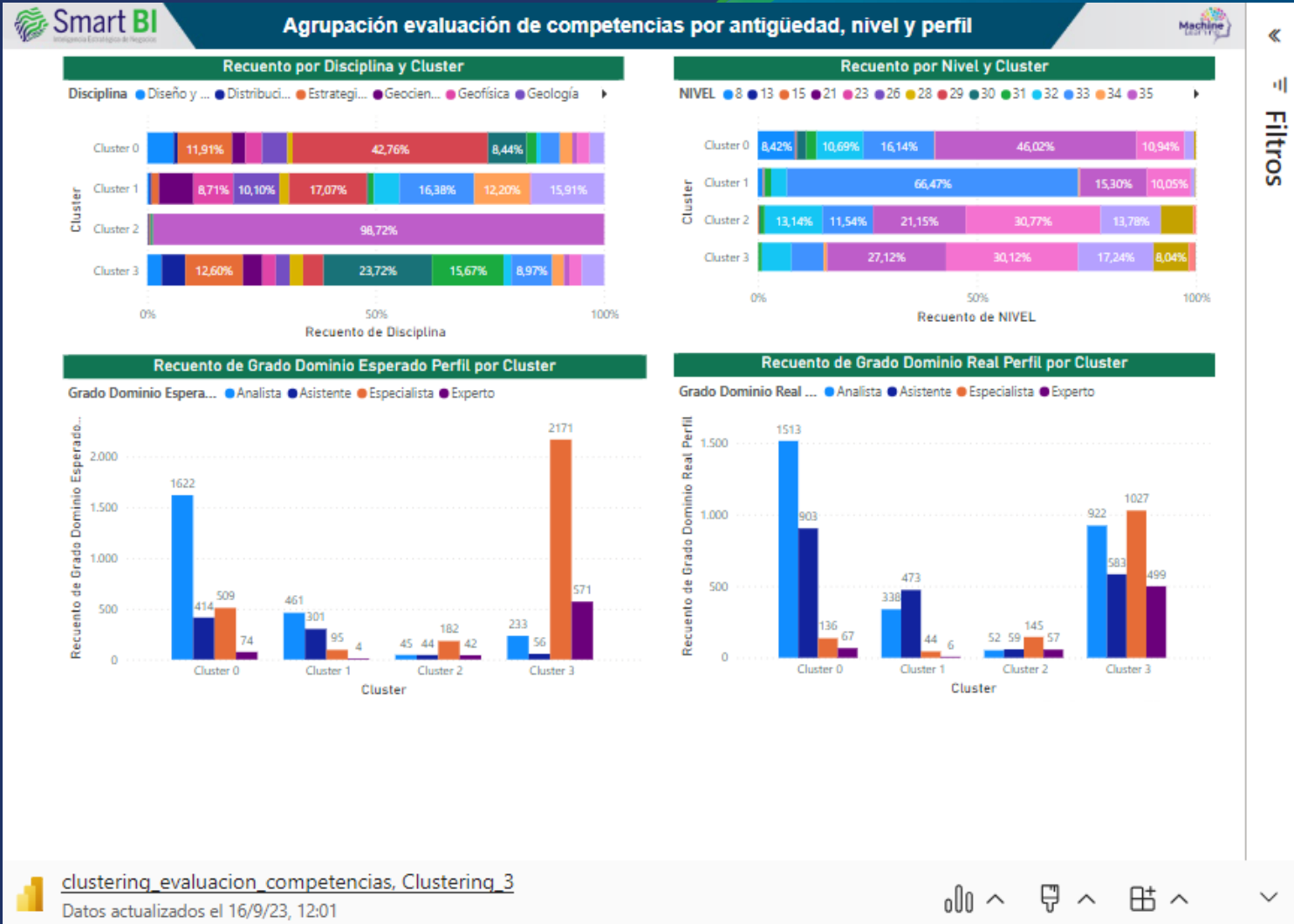
Densidad Aparente	Profundidad	Porosidad Efectiva	Porosidad Total	Gamma Ray	Resistividad	Tiempo de Viaje de la Onda de Corte	Tiempo de Viaje de la Onda de Compresión Real	Tiempo de Viaje de la Onda de Compresión Predicción
2,77	6805.0	0,00	0,03	110,48	57,47	124,05	67,68	70,70
2,77	6805.5	0,01	0,04	112,95	58,86	123,81	67,74	71,67
2,76	6806.0	0,01	0,04	114,49	65,02	123,41	67,75	71,76
2,75	6806.5	0,01	0,03	116,07	63,35	123,24	67,72	70,51
2,76	6807.0	0,00	0,03	117,68	62,28	123,44	67,69	70,56
2,77	6807.5	0,00	0,04	120,28	62,49	123,86	67,72	72,02
2,77	6808.0	0,00	0,04	121,49	62,62	124,41	67,85	73,26
2,77	6808.5	0,00	0,04	119,80	61,65	125,11	68,11	73,99
2,78	6809.0	0,01	0,04	119,04	62,96	125,85	68,38	74,73
2,78	6809.5	0,01	0,05	121,62	60,61	126,62	68,50	75,45
2,77	6810.0	0,01	0,05	122,70	57,18	127,36	68,48	75,86
2,76	6810.5	0,01	0,05	117,99	56,30	127,87	68,52	75,57
2,76	6811.0	0,01	0,05	114,28	56,99	127,92	68,58	75,32
2,76	6811.5	0,01	0,05	113,42	57,83	127,47	68,57	74,64
2,76	6812.0	0,01	0,04	114,13	57,16	126,88	68,54	72,99
2,74	6812.5	0,01	0,04	113,61	58,18	126,39	68,55	71,30
2,73	6813.0	0,01	0,04	114,43	59,84	126,16	68,63	70,74
2,75	6813.5	0,01	0,04	117,26	61,29	126,33	68,79	71,45
2,77	6814.0	0,00	0,04	118,54	58,41	126,81	69,02	72,85
2,79	6814.5	0,00	0,04	118,90	52,97	127,31	69,25	73,89

regresion propiedades geomecanicas... , predicción de modelo
Datos actualizados el 21/9/23, 12:39

Filtros

Análisis de Clustering – Evaluación de competencias.

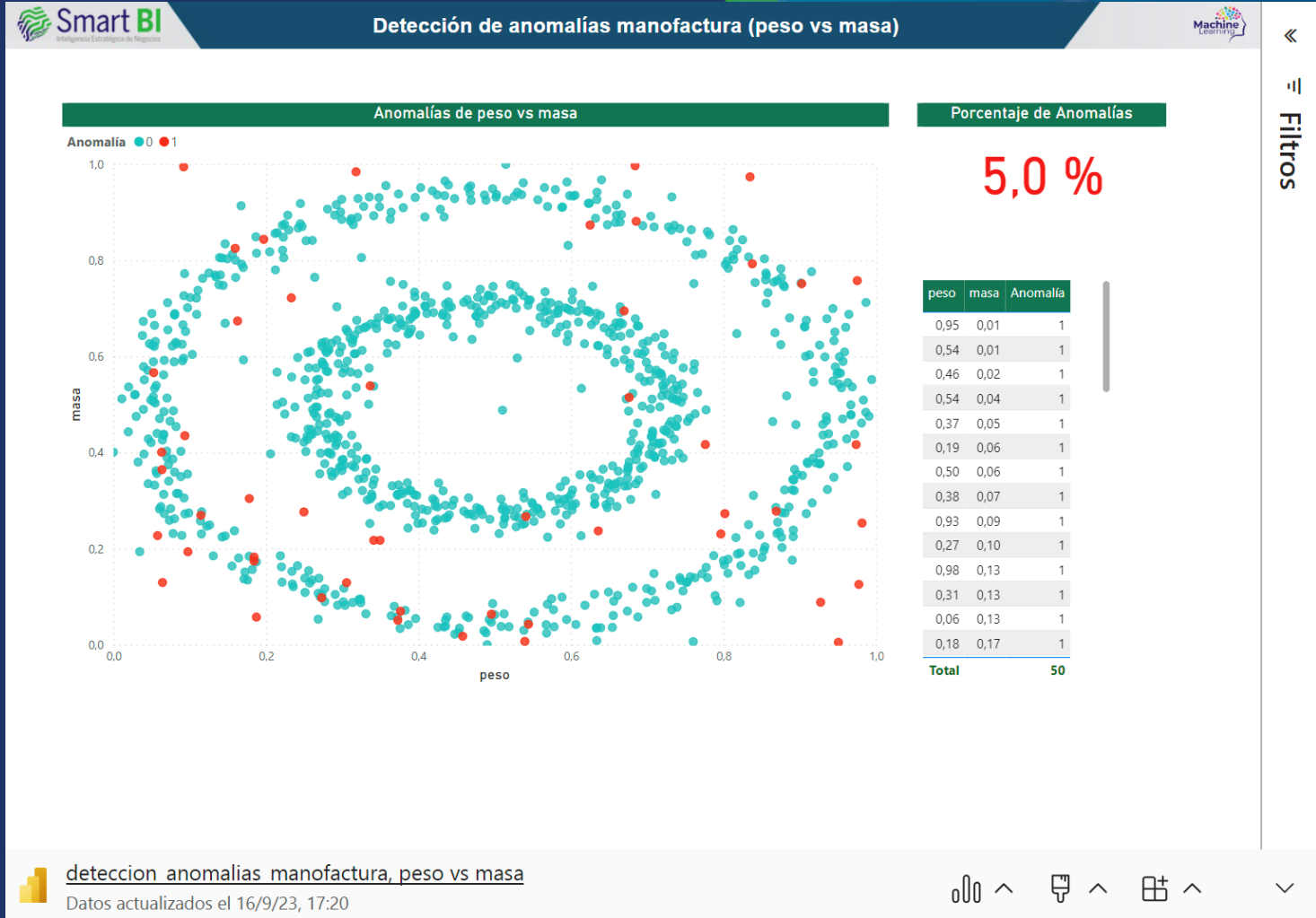
Su propósito es agrupar valores con características similares, en el ejemplo se muestra una agrupación de profesionistas con características similares. Hacemos uso del modelo kmeans de la librería Pycaret de Python para obtener las agrupaciones.



Análisis de Anomalías – Manofacura.

Su propósito es encontrar valores atípicos en un conjunto de datos, en el ejemplo se muestra un diagrama de dispersión entre el peso y la masa de un producto, en el cual podemos ver los valores atípicos en color rojo.

Hacemos uso del modelo anomaly de la librería Pycaret de Python para obtener las agrupaciones.





Smart BI

Inteligencia Estratégica de Negocios

Oficina. 52 (993) 1872396

Cel. 52 (993) 1609825

Correo: abocardog@smart-bi.mx

www.smart-bi.mx

