

CONCURSO DATOS AL ECOSISTEMA 2025 – RETO AGROSAVIA

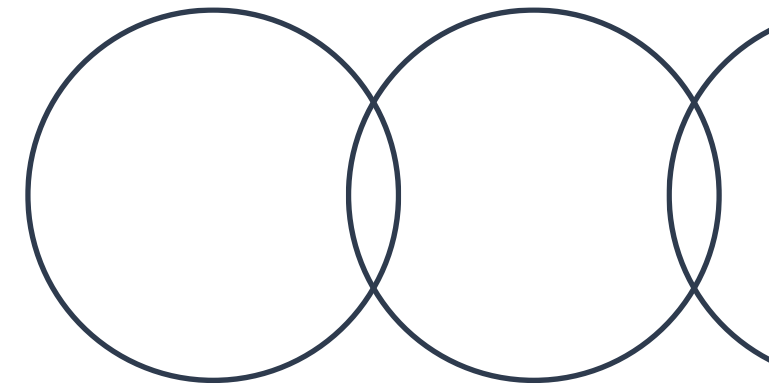
Herramienta Digital para Análisis de Suelos

Alejandro Garcia, Mauricio Zafra, Nelson
Galvis y Natalia Cuellar



Introducción a la herramienta

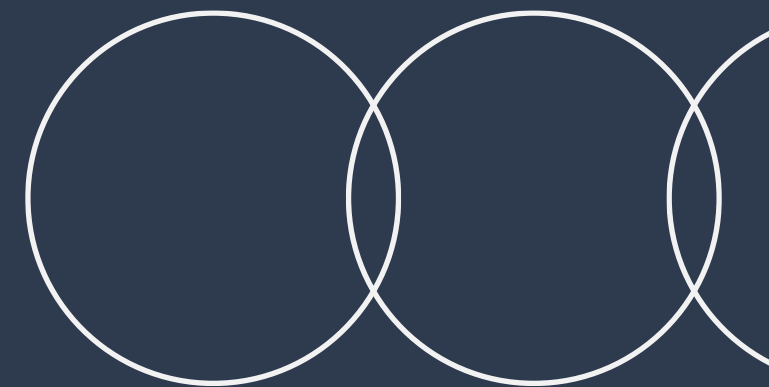
Con base en el reto planteado por agrosavia se diseño una herramienta que permitirá a los usuarios del conjunto de datos abiertos **Resultados de Análisis de Laboratorio Suelos en Colombia**, detectar ágilmente datos o registros anómalos, facilitar los análisis y evaluar la calidad de los datos suministrados.



Detección de Anomalías

Problemas en la Calidad

Actualmente, los usuarios del conjunto de datos de Agrosavia enfrentan desafíos para identificar **anomalías en los registros**, comprometiendo la calidad y confiabilidad de su análisis y decisiones informadas.



Estructura de los datos

secuencial	fecha_de_an_lisis	departamento	municipio	cultivo	estado	tiempo_de_establecimiento	topografia	drenaje	riego	...	conductividad_electrica	hierro_disponible_olsen	cobre_disponible	manganeso_disponible_olsen	zinc_disponible_olsen	boro_disponible	hierro_disponible_doble_acido	cobre_disponible_doble_acido	manganeso_disponible_doble_acido	zinc_disponible_doble_acido	
0	1	7/01/2014	NARIÑO	SAN ANDRES DE TUMACO	No Indica	No indica	No indica	No indica	No indica	...	0.133	66.39	1.700	1.700	1.5	0.187	ND	ND	ND	ND	
1	2	21/08/2014	HUILA	SANTA MARIA	Granadilla	Establecido	De 1 a 5 años	Pendiente	Buen drenaje	No Tiene	...	2.749	250.9	3.7	29.4	28.70	2.084	ND	ND	ND	ND
2	3	22/08/2014	ANTIOQUIA	LIBORINA	Café	Por establecer	No indica	Pendiente	Mal drenaje	No Tiene	...	0.328	390	11.10	8.299	5	0.085	ND	ND	ND	ND
3	4	22/08/2014	ANTIOQUIA	LIBORINA	Maracuyá	Por establecer	No indica	Ondulado	Mal drenaje	No Tiene	...	0.171	200	4.2	4.699	1	0.229	ND	ND	ND	ND
4	5	22/08/2014	ANTIOQUIA	LIBORINA	Café	Por establecer	No indica	Ondulado	Mal drenaje	No Tiene	...	0.323	117.9	3.8	4.600	0.700	0.139	ND	ND	ND	ND
5	6	22/08/2014	ANTIOQUIA	LIBORINA	Café	Por establecer	No indica	Pendiente	Mal drenaje	No Tiene	...	0.621	60.19	3.5	4.3	4.8	0.355	ND	ND	ND	ND
6	7	22/08/2014	ANTIOQUIA	LIBORINA	Café	Por establecer	No indica	Pendiente	Mal drenaje	No Tiene	...	0.184	330	5.5	6.4	1.400	0.175	ND	ND	ND	ND
7	8	22/08/2014	CUNDINAMARCA	SILVANIA	Mora	Establecido	De 5 a 10 años	Ondulado	Regular drenaje	No Tiene	...	0.709	535	9.200	3.400	4.699	0.247	ND	ND	ND	ND
8	9	22/08/2014	ANTIOQUIA	LIBORINA	Café	Por establecer	No indica	Ondulado	Mal drenaje	No Tiene	...	0.159	40.4	4.099	3.8	0.6	0.139	ND	ND	ND	ND
9	10	22/08/2014	CUNDINAMARCA	PASCA	Mora	Establecido	De 1 a 5 años	Pendiente	Regular drenaje	No Tiene	...	0.449	348	6.3	5.5	21.6	0.157	ND	ND	ND	ND

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 92738 entries, 0 to 92737
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   secuencial                            92738 non-null  object
1   fecha_de_an_lisis                     92738 non-null  object
2   departamento                         92738 non-null  object
3   municipio                             92736 non-null  object
4   cultivo                              92738 non-null  object
5   estado                               92738 non-null  object
6   tiempo_de_establecimiento             92738 non-null  object
7   topografia                           92738 non-null  object
8   drenaje                              92738 non-null  object
9   riego                                92738 non-null  object
10  fertilizantes_aplicados                92738 non-null  object
11  ph_agua_suelo                         92738 non-null  object
12  materiaorganica                       92738 non-null  object
13  fosforo_bray_ii                       92738 non-null  object
14  azufre_fosfato_monocalcico            92738 non-null  object
15  acidez_kcl                            92738 non-null  object
16  aluminio_intercambiable               92738 non-null  object
17  calcio_intercambiable                 92738 non-null  object
18  magnesio_intercambiable               92738 non-null  object
19  potasio_intercambiable                92738 non-null  object
20  sodio_intercambiable                  92738 non-null  object
21  capacidad_de_intercambio_cationico    92738 non-null  object
22  conductividad_electrica               92738 non-null  object
23  hierro_disponible_olsen                92738 non-null  object
24  cobre_disponible                      92738 non-null  object
25  manganeso_disponible_olsen            92738 non-null  object
26  zinc_disponible_olsen                 92738 non-null  object
27  boro_disponible                       92738 non-null  object
28  hierro_disponible_doble_acido          92737 non-null  object
29  cobre_disponible_doble_acido           92738 non-null  object
30  manganeso_disponible_doble_acido       92738 non-null  object
31  zinc_disponible_doble_acido            92738 non-null  object
dtypes: object(32)
memory usage: 22.6+ MB
```

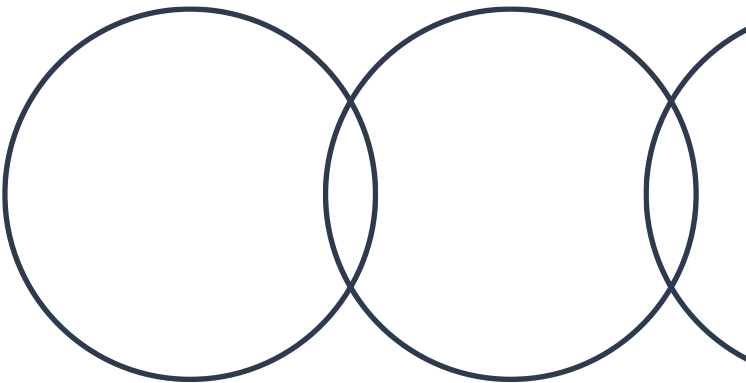
92738 filas y 32 columnas

Datos registrados como object

Diferentes subgrupos por digitaciones

Columnas con más de 90% de datos ND

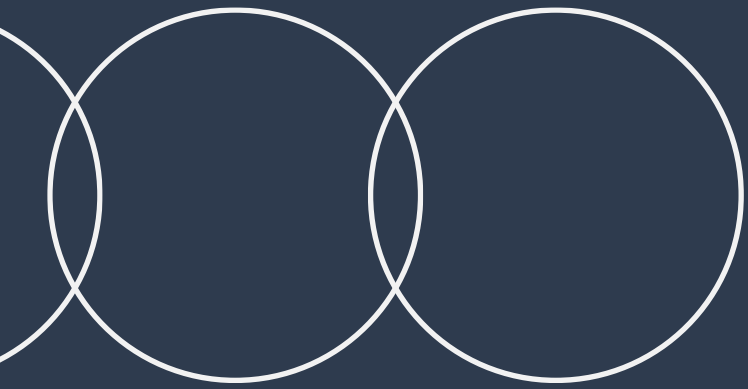
Exigió un pretaratamiento de datos más riguroso sin comprometer la estructura ni la calidad de los datos



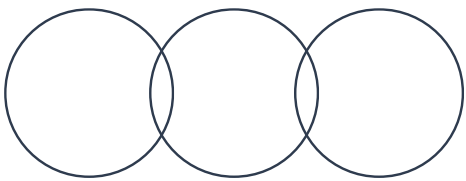
Innovación y Creatividad

Enfoque Creativo

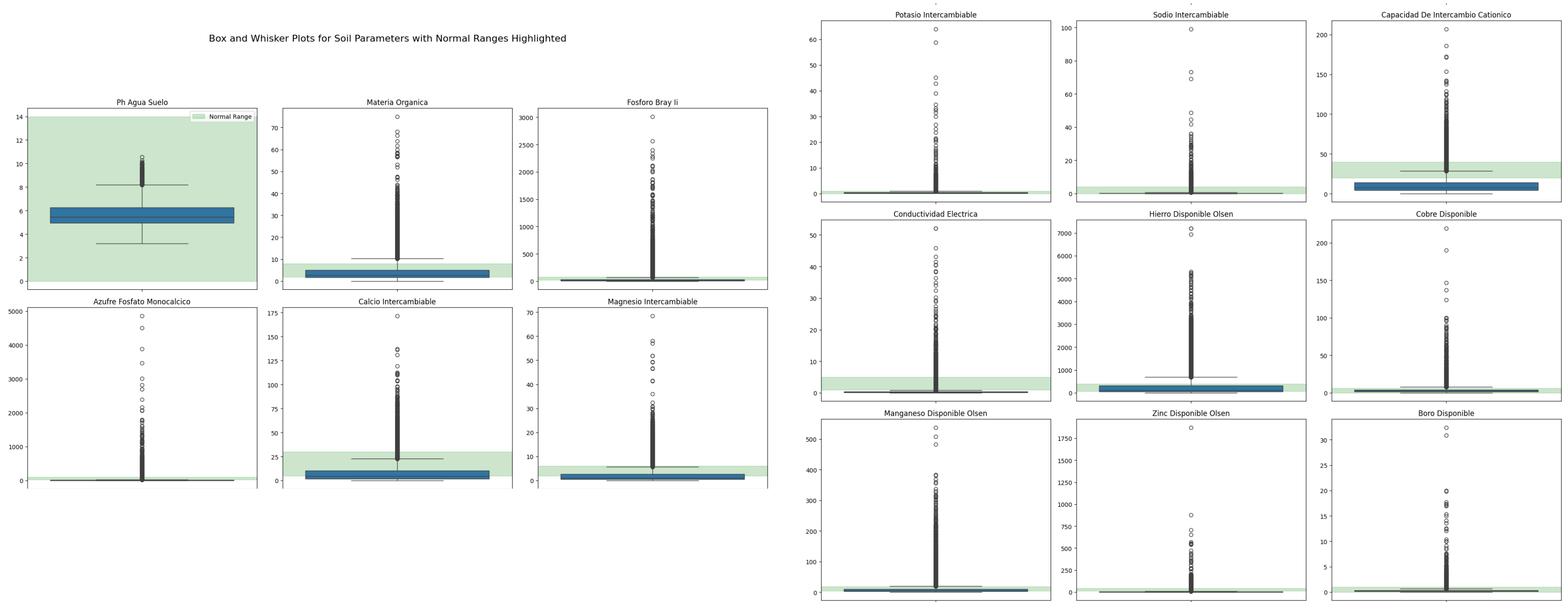
La herramienta digital se inspira en **soluciones innovadoras**, utilizando técnicas avanzadas para detectar anomalías, mejorando así la experiencia del usuario y la efectividad del análisis de datos.



Interfaz y Funcionalidades

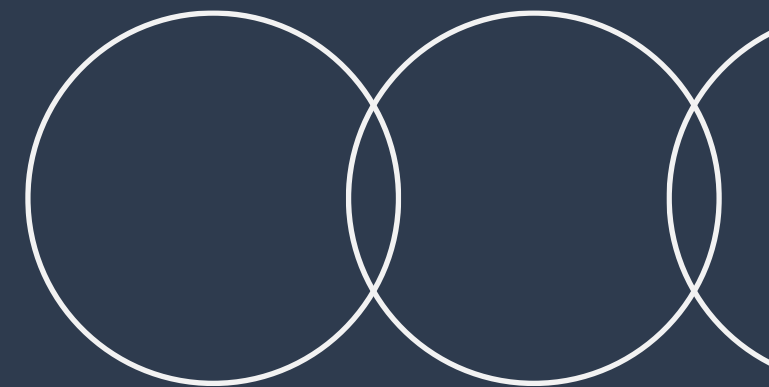


Box and Whisker Plots for Soil Parameters with Normal Ranges Highlighted



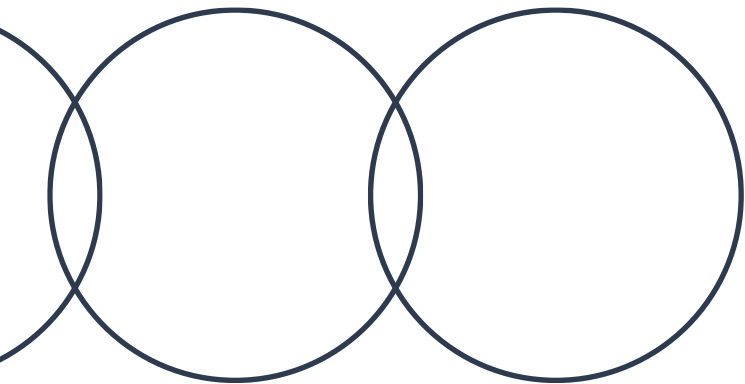
Impacto Esperado

La herramienta permite evaluar la calidad de los datos que se vinculan a cada una de las variables de análisis de suelos e indicar la calidad del mismo. De esa manera y con base a criterios agronómicos podría facilitar la interpretación para los diferentes tipos de suelo y de cultivo y si es o no necesario el uso de programas de fertilización o modificadores del suelo.



Análisis Técnico

Detección de anomalías en registros de suelo



Análisis Rigoroso

Se realizan análisis estadísticos avanzados para identificar **anomalías en los datos**, asegurando un resultado confiable y preciso.

Métodos Estadísticos

Empleamos algoritmos específicos y técnicas de **machine learning** que permiten detectar patrones irregulares en los registros de análisis de suelo.

Tecnologías Emergentes

Incorporamos herramientas de **inteligencia artificial** y big data para gestionar datos anómalos con alta precisión y eficiencia.

Ejemplos Técnicos de Análisis

Análisis de datos Hallazgos clave

Éxito en el entrenamiento del modelo: Los modelos regresores de bosque aleatorio se entrenaron con éxito para las 12 columnas numéricas elegibles tras resolver varios problemas técnicos, incluyendo la gestión de dependencias, la instalación de módulos y, lo más importante, asegurar la conversión correcta de tipos de datos a float64 para columnas numéricas.

Previsibilidad variable - Puntuaciones R2:

Altamente predecible (mostró una muy fuerte previsibilidad por otros parámetros del suelo)

calcio_intercambiable (R2: 0,988)

capacidad_de_intercambio_cationico (R2: 0,984)

Moderadamente predecible (mostró una previsibilidad razonable):

ph_agua_suelo (R2: 0,875)

magnesio_intercambiable (R2: 0,925)

potasio_intercambiable (R2: 0,659)

conductividad_electrica (R2: 0,704)

Menos predecibles (eran menos predecibles, lo que indica que su variabilidad no se explica tan bien por el conjunto actual de características)

materiaorganica (R2: 0,501)

azufre_fosfato_monocalcico (R2: 0,487)

hierro_disponible_olsen (R2: 0,527)

fosforo_bray_ii (R2: 0,338)

manganeso_disponible_olsen (R2: 0,323)

boro_disponible (R2: 0,339)

Valores de Error Cuadrático Medio Raíz

(RMSE): Los valores RMSE variaron

significativamente entre modelos,

reflejando la escala de las variables

objetivo. Por ejemplo,

calcio_intercambiable tuvo un RMSE

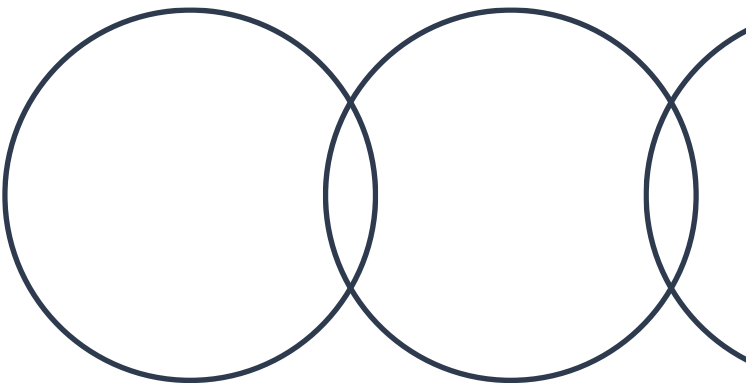
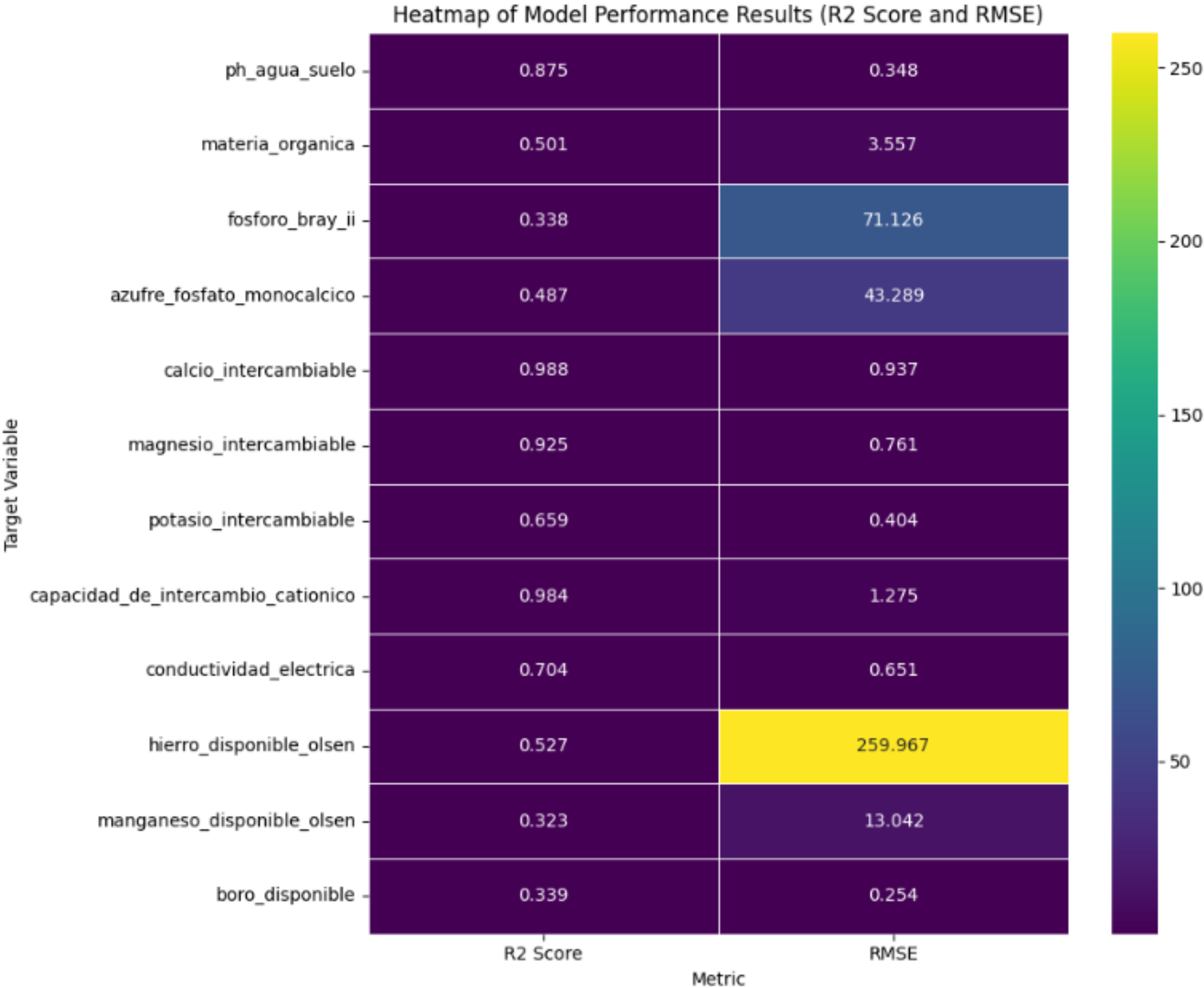
mínimo de 0,937, mientras que

hierro_disponible_olsen tuvo un RMSE

sustancialmente mayor de 259,967.



Impacto en Calidad



Escalabilidad de la Herramienta



01

La herramienta puede adaptarse a diferentes conjuntos de datos fácilmente.



02

Posible expansión a las regiones agrícolas y cultivos con similares características.

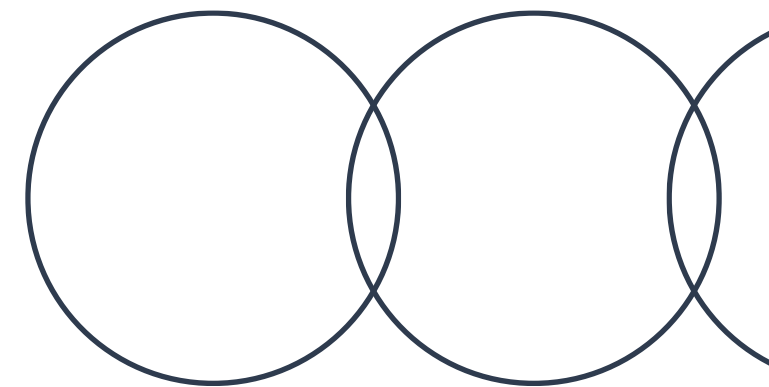


03

Interfaz flexible permite personalización según requerimientos específicos del usuario.

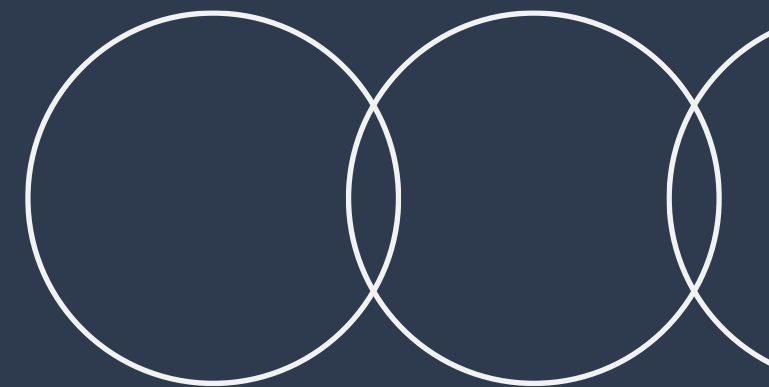
Diseño y Usabilidad

Aun se deben mejorar varios detalles del diseño y usabilidad



Conclusión y Próximos Pasos

Fue posible generar el indicador de calidad de datos. Sin embargo, considerando la variabilidad temporal del conjuntos de datos y todas las posibles variables metodológicas que pueden intervenir en el análisis, se recomienda que se favorezcan subconjuntos de datos como: servicio o proyecto, tipo de muestreo realizado, requerimiento de muestreo. Quizás de esta forma pueda darse una mejor estructura y análisis al compendio de datos.



¿Listo para mejorar los análisis?

En busca de mejorar la agricultura en Colombia

