

Ciencia de datos al servicio del petróleo: eligiendo el mejor levantamiento artificial

PROBLEMA



La selección del sistema de levantamiento artificial en el campo Lago Agrio se realiza con criterios no estandarizados y registros históricos dispersos (BPPD, BFPD, BSW, API, GOR, TVD) que no se integran de forma sistemática. Esto deriva en decisiones inconsistentes, desajustes pozo-sistema, menor eficiencia productiva y mayores costos de operación e intervención.



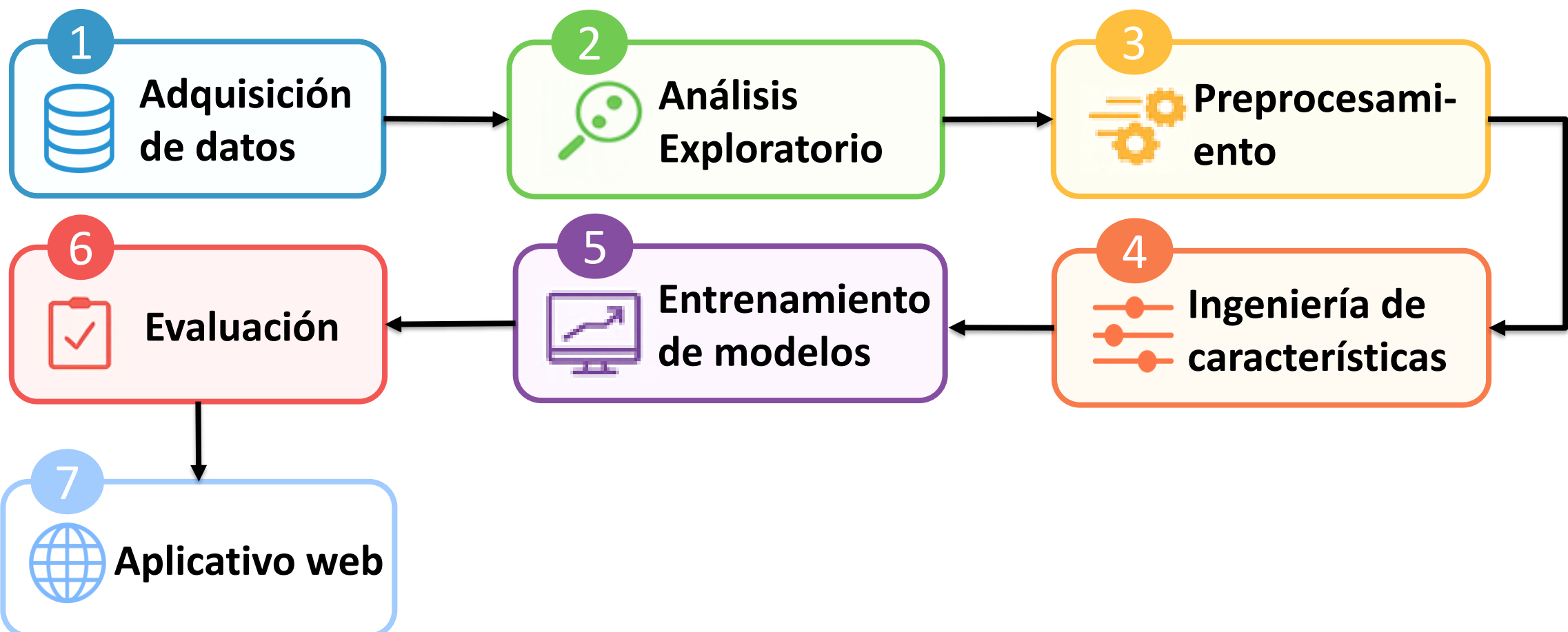
OBJETIVO GENERAL



Proponer sistemas de levantamiento artificial más adecuados para pozos del campo Lago Agrio mediante ciencia de datos, optimizando la producción del activo hidrocarburífero.



PROPUESTA



Adquisición de datos: Se consolidaron historiales por pozo y se verificó consistencia y trazabilidad de fuentes y eventos

Análisis exploratorio: Exploración de distribuciones estadísticas de las variables analizadas y detección de outliers.

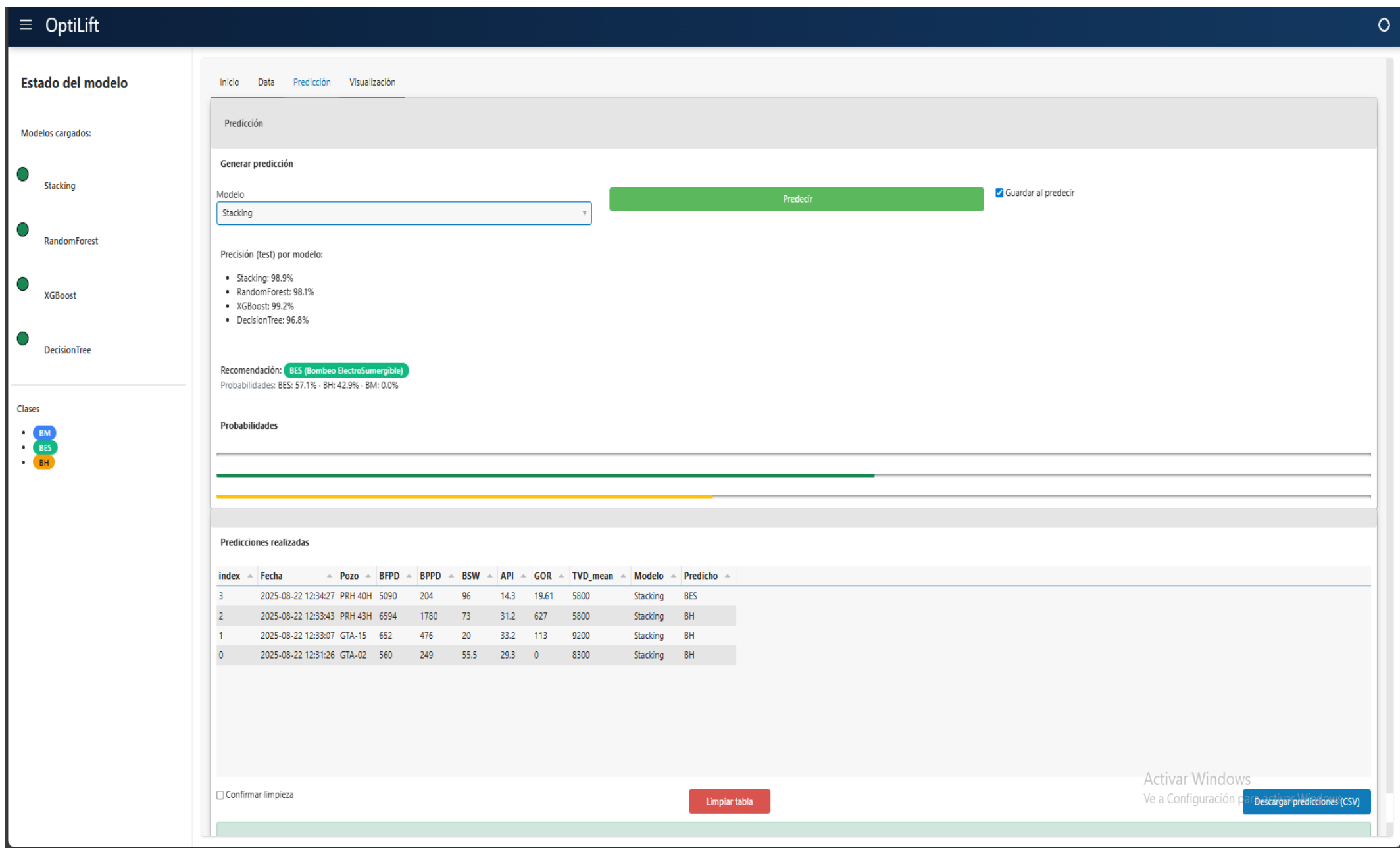
Preprocesamiento: Limpieza de datos, tratamiento de datos faltantes/atípicos, estandarización de características.

Ingeniería de características: Generación y selección de características (BFPD, BPPD, BSW, API, GOR, TVD_mean)

Entrenamiento de modelos: Se entrenaron modelos con aplicación de validación cruzada y ajuste de hiperparámetros

Evaluación: Se compararon métricas y se eligió mejor modelo por desempeño y robustez

Aplicativo Web: Se integró los modelos para la generación de recomendación de SLA por pozo



RESULTADOS

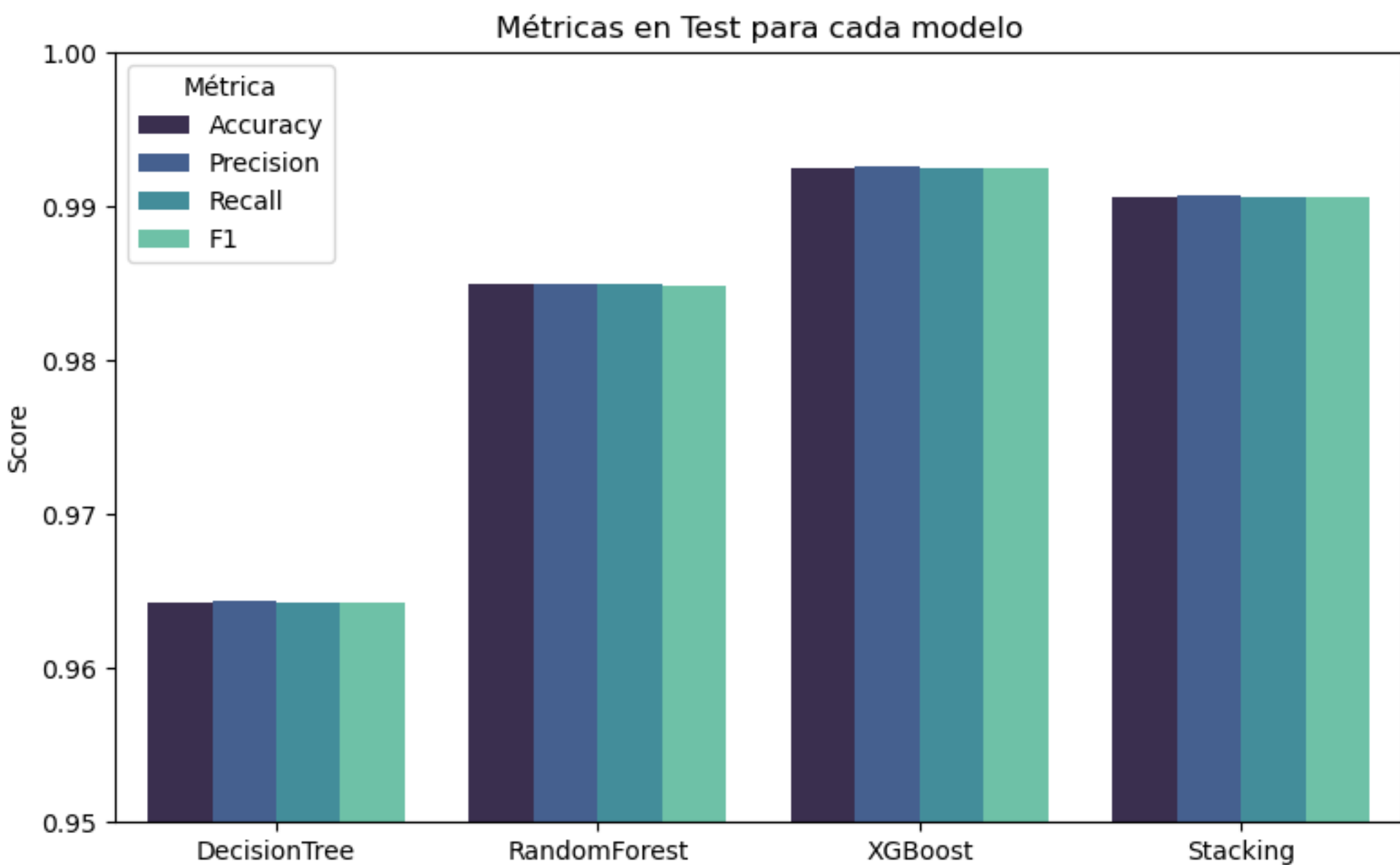


Predicción de SLA por pozo (Stacking)

Pozo	BFPD	BPPD	BSW (%)	API	GOR (scf/bbl)	TVD_mean (ft)	Predicho
PRH 40H	5,090	204	96.0	14.3	19.61	5,800	BES
PRH 43H	6,594	1,780	73.0	31.2	627.0	5,800	BH
GTA-15	652	476	20.0	33.2	113.0	9,200	BH
GTA-02	560	249	55.5	29.3	0.0	8,300	BH

Comparación de métricas de modelos

Modelo	Exactitud	Precisión	Recall	F1-Score
Árbol de decisión	0.964	0.964	0.964	0.964
Random Forest	0.960	0.960	0.960	0.960
XGBoost	0.992	0.992	0.992	0.992
Stacking Classifier	0.991	0.991	0.991	0.991



CONCLUSIONES



Se evidenció que todos los modelos ensayados alcanzaron exactitudes superiores al 96%. En particular, XGBoost y Stacking Classifier se aproximaron al 99%, lo que confirmó la eficacia de los métodos de *ensemble* para este tipo de clasificación y su utilidad para apoyar la selección de sistemas de levantamiento artificial en campos maduros.



El uso de ciencia de datos transformó la forma de seleccionar sistemas de levantamiento artificial, pasando de un enfoque tradicional a uno basado en predicciones basadas en datos técnicos. Esta metodología aporta una herramienta práctica para la industria petrolera, capaz de mejorar la toma de decisiones.

