

## ARTÍCULOS

# Predicción de Producción de Fluidos empleando Machine Learning en T Inferior del Campo Sacha



## Fluid Production Prediction using Machine Learning in the Lower T of the Sacha Field

Altamirano-Cárdenas, Adrián Israel; Lucero-Calvache, Fernando Andrés

 **Adrián Israel Altamirano Cárdenas**  
aialtamiranoc@uce.edu.ec  
Universidad Central del Ecuador. Carrera de Ingeniería de Petróleos. Quito, Pichincha, Ecuador.

 **Fernando Andrés Lucero Calvache**  
falucero@uce.edu.ec  
Universidad Central del Ecuador. Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental. Quito, Pichincha, Ecuador.

### FIGEMPA: Investigación y Desarrollo

Universidad Central del Ecuador, Ecuador  
ISSN: 1390-7042  
ISSN-e: 2602-8484  
Periodicidad: Semestral  
vol. 16, núm. 2, 2023  
[revista.figempa@uce.edu.ec](mailto:revista.figempa@uce.edu.ec)

Recepción: 04 abril 2023  
Aprobación: 30 junio 2023

DOI: <https://doi.org/10.29166/revfig.v16i2.4542>

Autor de correspondencia: [aialtamiranoc@uce.edu.ec](mailto:aialtamiranoc@uce.edu.ec)



Esta obra está bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

**Cómo citar:** Altamirano-Cárdenas, A. I.; Lucero-Calvache, F. A. (2023). Predicción de Producción de Fluidos empleando Machine Learning en T Inferior del Campo Sacha. *FIGEMPA: Investigación y Desarrollo*, 16(2), 70-78. <https://doi.org/10.29166/revfig.v16i2.4542>

**Resumen:** El presente estudio tiene por objeto predecir la producción de fluidos de los pozos perforados durante el año 2021 en la arenisca T inferior en el campo Sacha. Se tomó en cuenta la información petrofísica y de fluidos para la construcción del modelo a partir de un pozo tipo mediante el software comercial el cual proporciona el análisis del comportamiento de pozos. Paralelamente mediante el lenguaje de programación Python a través de Machine Learning se desarrolló dos algoritmos: uno basado en los datos de presión de entrada (PIP) de la bomba electrosumergible (BES), y otro con datos PIP y salinidad del agua de formación del reservorio. La predicción de producción de fluidos respecto a la producción real obtuvo un error de 2% con el software comercial mientras que en las dos simulaciones a través de Python se obtuvo 10% y 0.5% de error respectivamente. Para el caso de la predicción de producción de gas el valor real es de 0.07 MMSCFD mientras que el obtenido mediante la simulación con el software comercial es de 0.41 MMSCFD. Para el caso de la primera y segunda simulación con Python se obtuvo una mejor aproximación de 0.11 MMSCFD y 0.10 MMSCFD respectivamente. El incremento de variables en Python permite la reducción del porcentaje de error e incrementa el ajuste de la predicción de producción de fluidos y gas, en este caso la PIP de la BES y salinidad de agua de formación.

**Palabras clave:** análisis de datos; pronóstico de producción; software de código abierto; tecnología petrolera.

**Abstract:** The purpose of this study is to predict the fluid production of the wells drilled during the year 2021 in the lower T sandstone in the Sacha field. The petrophysical and fluid information was considered for the construction of the model from a standard well using commercial software, which provides the analysis of well behavior. In parallel, using the Python programming language through Machine Learning, two algorithms were developed: one based on the pump intake pressure (PIP) data of the electric submersible pump (ESP), and another with PIP data and salinity of the reservoir formation water. The prediction of fluid production with respect to the real production, found an error of 2% with the

commercial software while in the two simulations through Python an error of 10% and 0.5% was obtained respectively. In the case of gas production prediction, the real value is 0.07 MMSCFD, while the one obtained by simulation with commercial software is 0.41 MMSCFD. For the case of the first and second simulation with Python, a better approximation of 0.11 MMSCFD and 0.10 MMSCFD was obtained respectively. The increase of variables in Python allows the reduction of the percentage of error and increases the adjustment of the of fluid and gas production prediction, in this case the PIP of the BES and salinity of formation water.

**Keywords:** data analysis; production forecasting; open-source software; petroleum technology.

## INTRODUCCIÓN

El objetivo final del modelamiento de un campo petrolero es idear una estrategia óptima para desarrollar, administrar y operar el campo. Para algunos campos petroleros, la optimización de las operaciones de producción puede ser un factor importante para aumentar las tasas de producción y reducir los costos. Mientras que para pozos individuales u otros sistemas pequeños el análisis nodal simple puede ser adecuado, los sistemas grandes y complejos exigen un enfoque mucho más sofisticado para predecir con precisión la respuesta de un sistema de producción y para examinar escenarios operativos alternativos de manera eficiente.

A medida que los algoritmos de optimización y las técnicas de simulación de yacimientos continúan desarrollándose y la capacidad computacional continúa aumentando, las instalaciones de petróleo y gas que antes se pensaba que no eran candidatas para el control u optimización avanzados están siendo objeto de nuevas consideraciones.

El presente estudio compara las predicciones de producción de un pozo tipo de la arenisca T inferior del campo Sacha con las obtenidas mediante software comercial con el fin de mejorar el método de cálculo y tener mayor exactitud del volumen de fluidos a producir para que las facilidades de producción y de levantamiento de los fluidos sean las idóneas. Además, se presenta propuestas en base a los resultados obtenidos, los cuales servirán como fundamentos teóricos y técnicos para futuros proyectos de explotación de pozos.

## MATERIALES Y MÉTODOS

El presente estudio se ha aplicado en un enfoque analítico - descriptivo, en el cual se analizó la predicción de la producción de agua, gas y petróleo, a través de la utilización de software comercial en la arena T inferior del campo Sacha basados en los datos de presión de entrada de la bomba electrosumergible y la producción de los pozos del año 2021. En la Tabla 1 se presentan los datos de producción del pozo tipo.

**TABLA 1**  
Datos de producción del pozo tipo

<b>Producción fluido total (BFPD)</b>	<b>Producción de gas (MMSCFD)</b>	<b>Producción de petróleo (BPPD)</b>	<b>Producción de agua (BWPD)</b>	<b>BSW (Fracción)</b>	<b>GOR (SCF/STB)</b>
525	0.07	498.75	26.25	0.05	800

## Consideraciones teóricas

Los métodos de optimización de machine learning en la industria petrolera se introdujeron por primera vez en la industria de exploración y producción (E&P) de petróleo en la década de 1940, y desde entonces se han utilizado ampliamente para predecir, estimar y determinar varios parámetros operativos. Estos métodos se clasifican principalmente en tres grupos que incluyen técnicas de programación lineal, entera y no lineal.

Los algoritmos de inteligencia artificial (IA) se han utilizado ampliamente para resolver problemas en la industria del petróleo y el gas. Zhou (2019) menciona que la IA se define como la capacidad de los agentes inteligentes para el aprendizaje continuo en el entorno correspondiente y la percepción de determinadas actividades. AI se compone principalmente de Algoritmos Evolutivos (EA), Inteligencia de Enjambre (SI), Lógica Difusa (FL) y Redes Neuronales Artificiales (RNA).

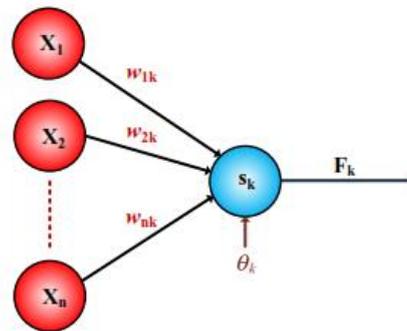
## Red neural artificial (RNA)

Consiste en un conjunto de unidades de procesamiento simples, que comunicarse mediante el envío de señales entre sí a través de un gran número de conexiones ponderadas. La neurona recibe la información de sus vecinos o fuentes externas y usar esta entrada para calcular una señal de salida que luego se propaga a otras unidades. Además de este procesamiento, la segunda tarea es el ajuste de los pesos. El sistema es inherentemente paralelo en el sentido de que muchas unidades pueden realizar sus cálculos al mismo tiempo. En la mayoría de los casos, se supone que cada unidad proporciona una contribución aditiva a la entrada de la unidad con la que está conectado (Rahmanifard & Plaksina, 2018).

Como se muestra en la Figura 1, la entrada total ( $sk(t)$ ) a la unidad  $k$  es una simple suma ponderada de las salidas de cada una de las unidades conectadas más un término de polarización o compensación ( $\theta_k$ ) por lo que se requiere una regla que calcule el efecto de la entrada total sobre la salida de la unidad.

Por lo tanto, una función ( $F_k$ ) que toma la entrada total  $sk(t)$  y produce un nuevo valor de la se necesita la activación de la unidad  $k$  (función de activación o transferencia). Las funciones de transferencia sirven principalmente como un tipo de filtro o puerta que permite que algunas señales se muevan avanzar y detener a otros a medida que avanzan desde los nodos de entrada a los de salida. lo más las funciones de activación comúnmente utilizadas son sigmoide logarítmica,

sigmoide tangente hiperbólica, y funciones lineales.

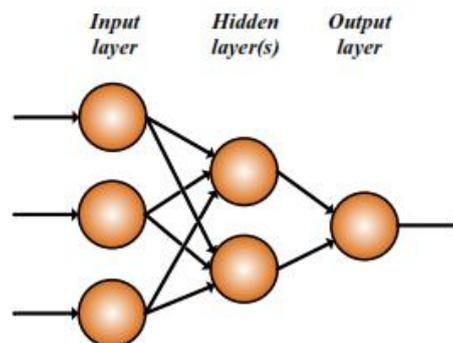


**FIGURA 1**

Ilustración de los conceptos básicos de una neurona

Fuente: Rahmanifard & Plaksina (2018)

Hay varios patrones de RNA, como redes feed-forward, redes recurrentes, etc. Las redes feed-forward convencionales son las redes más comunes para la aproximación de funciones mismas que constan de grupos de nodos interconectados dispuestos en capas correspondientes a entrada, oculta y nodos de salida, se muestra en la Figura 2.



**FIGURA 2**

Red Neural Artificial

Fuente: Rahmanifard & Plaksina (2018)

## Lenguaje Python

El lenguaje de programación Python es un lenguaje fácil de aprender. Éste cuenta con estructuras de alto nivel que puede dar un enfoque simple pero asertivo hacia la programación dirigida a objetos. Las características que se revelan en su sintaxis y su tipado dinámico lo transforman en un lenguaje ideal para scripting y para el uso de aplicaciones en diferentes áreas (inteligencia artificial, big data, data science, desarrollo web, entre otros) (Ibrahim & Bilchick, 2021). En la Figura 3 se presenta el algoritmo utilizado en el lenguaje de programación Python para las simulaciones.

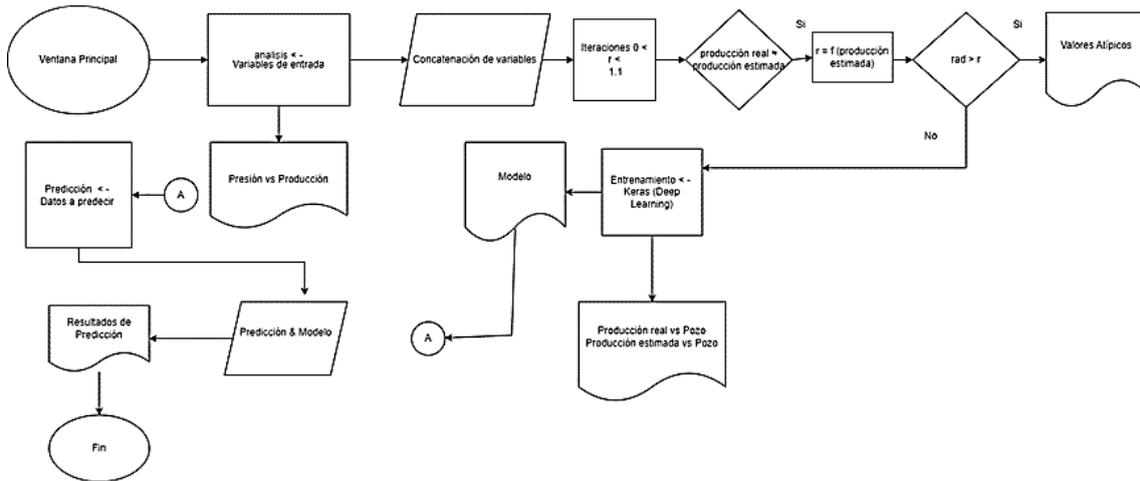


FIGURA 3

Algoritmo lenguaje de programación Python

### Análisis y depuración de datos

Para la simulación fue utilizada la normalización de datos para el posterior uso de clustering con el objetivo de eliminar los valores atípicos. Se realizó un cálculo iterativo de radio vs producción con el fin de obtener el mejor ajuste. Para el caso de estudio se determinó si los datos tienen un rango mayor al radio de 0.2 para fluidos y 0.8 para gas es un valor atípico. La Figura 4 muestra la ventana de análisis y depuración de datos donde se grafican la PIP vs producción de fluido (izquierdo) y la PIP vs producción de gas (derecho).

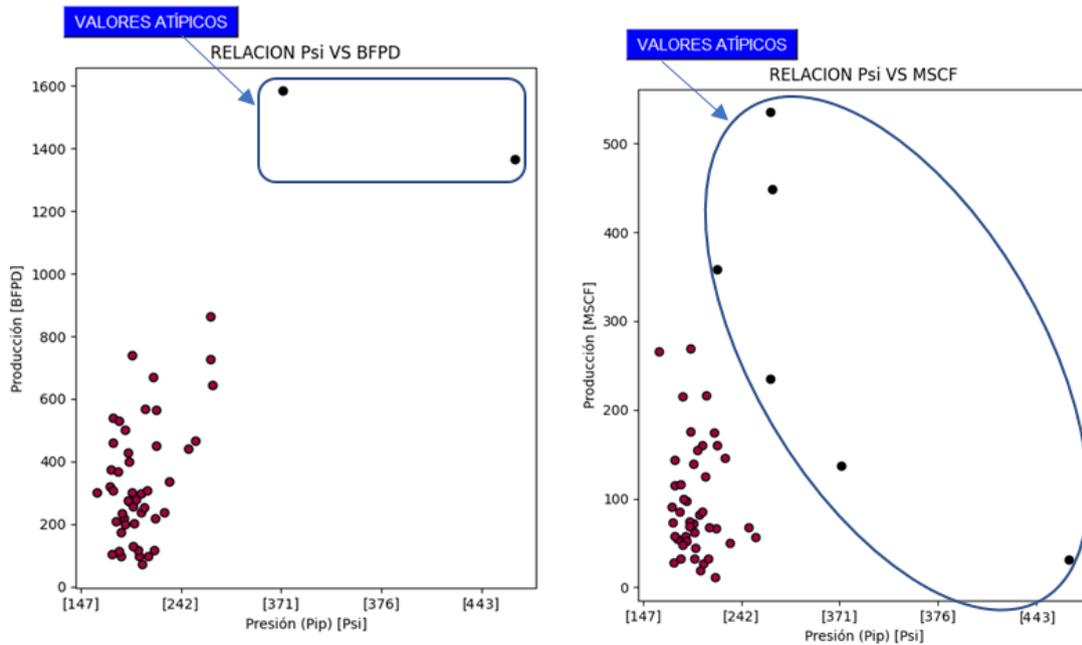


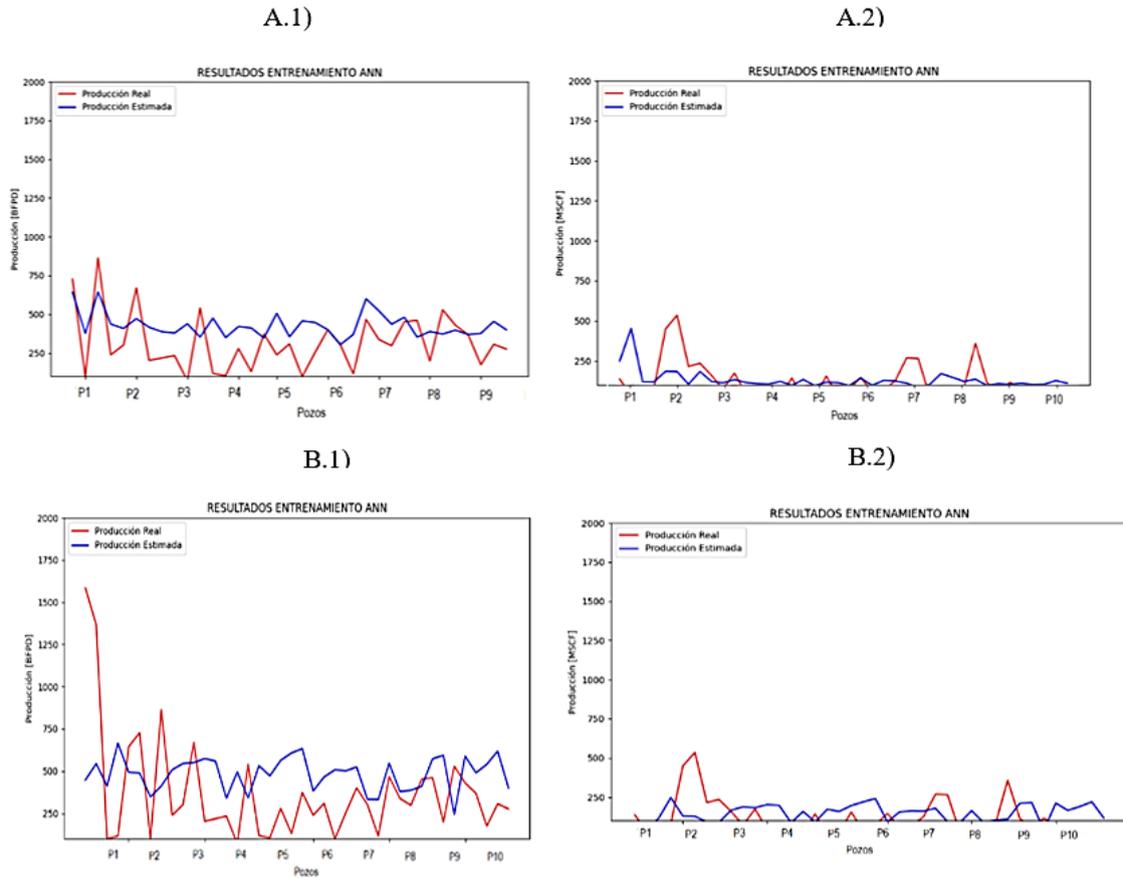
FIGURA 4

Análisis y depuración de valores atípicos de fluido y gas

Fuente: Software Python

## Modelo entrenado

Luego de la depuración de datos se realizó el entrenamiento de la RNA, en la que se obtuvo gráficas comparativas de la producción real y estimada con la red neuronal de los pozos como se observa en la Figura 5. La primera estimación se utilizó como variable la PIP, mientras que en la segunda estimación a más de la PIP se utilizó la salinidad del agua de formación. En la sobreposición de las curvas se observa el mejor ajuste del algoritmo que permitirá la predicción de la producción.



**FIGURA 5**

Modelo entrenamiento de la RNA.

A. Primera simulación: A.1) Producción de fluidos; b) Producción de gas; B. Segunda simulación: B.1) Producción fluidos; B.2) Producción de gas.

Fuente: Software Python

## RESULTADOS Y DISCUSIÓN

Para determinar la producción con el software comercial se utilizó la información disponible del pozo tipo. De forma paralela se realizaron dos predicciones de producción de fluido y gas: la primera con el valor de la PIP del pozo tipo de 250 Psi, y la segunda con el mismo valor de la PIP de la primera simulación y el valor de la salinidad del agua de formación de 50000 ppm. A partir de estos resultados se realizó la comparación de producción real versus la simulada tanto en el software comercial como en Python del pozo tipo como se indica en la Tabla 2.

**TABLA 2**

Comparativo de producción real y simulada en software comercial y Python del pozo tipo

Caso de estudio		Producción de fluido (BFPD)	Producción de Gas (MMSCFD)	Error producción de fluido (%)
<b>Producción real</b>		525.00	0.07	-----
<b>Software Python</b>	Primera simulación	473.40	0.11	9.83
	Segunda Simulación	522.44	0.10	0.49
<b>Software Comercial</b>		534.30	0.41	1.77

### Análisis comparativo de las simulaciones

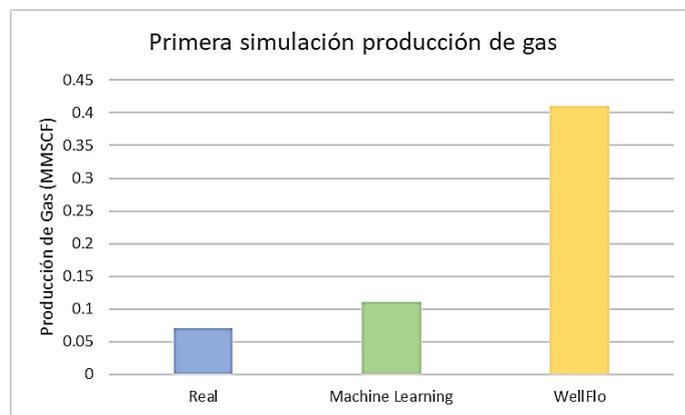
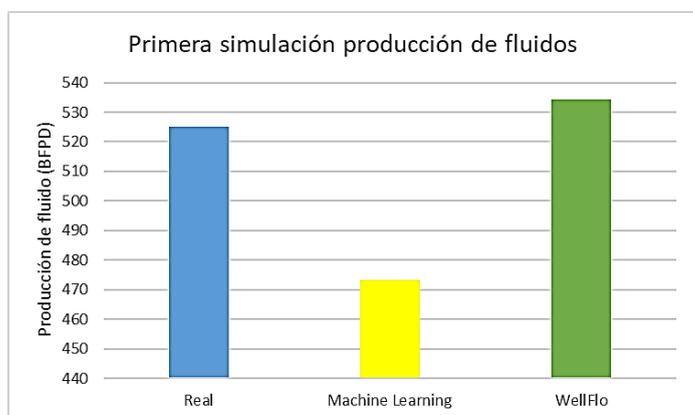
En la primera simulación se obtuvo un error del 10% con respecto al valor real de producción de fluidos del pozo tipo ya que únicamente se utilizó una variable. En la segunda simulación, el error se redujo a 0,5% debido a que se incrementó una variable adicional para obtener un mejor ajuste. Previamente se había realizado simulaciones con variables como la temperatura del pozo y el grado API donde no se determinó un cambio significativo de resultados por ser valores promedios para todos los pozos del Campo. Por el contrario, la PIP de la BES se relaciona directamente con la producción y la salinidad con la porosidad, permeabilidad, saturación de agua, arena productora y presión capilar, que son propiedades directamente relacionadas en el potencial de producción. En la Figura 6 se muestran los gráficos de barras de las simulaciones.

**FIGURA 6**

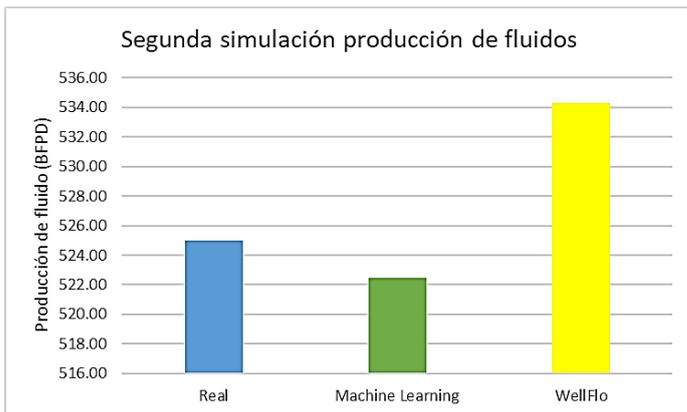
Gráfico de barras de las simulaciones

A.1)

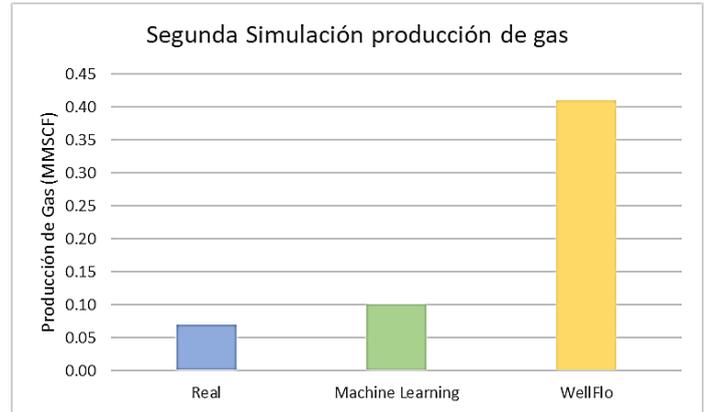
A.2)



B.1)



B.2)



A.1) producción de fluidos primera simulación, A.2) producción de gas primera simulación, B.1) producción de fluidos segunda simulación B.2) producción de gas segunda simulación

## CONCLUSIONES

- El uso de nuevas tecnologías como el machine learning permite que las predicciones de producción sean más precisas y se reduzca la incertidumbre.
- La concatenación y normalización de datos permiten eliminar los valores atípicos por lo cual es importante tener criterios adecuados para excluir estos valores como conocer el mejor método para este caso fue el clustering y dentro de este la variable radio con el mejor ajuste de 0.8 para gas y 0.2 para fluidos.
- La reutilización de código en las funciones optimiza espacio, y contribuye con la eficiencia del programa, esto puesto que tanto para gas como para fluidos se utiliza el mismo código además de que las funciones que requieran cálculos similares se utiliza la misma función en la que ya se realizó ese cálculo previo.
- Las variables que actúan directamente en el potencial de producción del pozo tipo reducen el porcentaje de error en la predicción de producción con respecto al software comercial.
- El mejoramiento de la predicción de producción se debe al uso de las redes neuronales artificiales que son las responsables de crear algoritmos de inteligencia artificial, estos permiten emular el comportamiento no lineal presentado por la PIP y la salinidad con respecto a la producción de fluido.

## REFERENCIAS

- Cass, S. (05 de Noviembre de 2019) *IEEE Spectrum*. Obtenido de IEEE Spectrum: <https://spectrum.ieee.org/at-work/innovation/the-2018-top-programminglanguages>.
- Flores Urgilés, C. M. y Ortiz Amoroso, M. S. (2018) Revisión de algoritmos para la detección de valores atípicos. *Killkana Técnica*, 2(1), pp. 19–26. doi: 10.26871/killkana\_tecnica.v2i1.287.
- Hunter, J., & Dale, D. (05 de Nov de 2019) *Matplotlib*. Obtenido de Python plotting: <https://matplotlib.org/>.

- Ibrahim, M., & Bilchick, K. (2021) *Avanzado Método de aprendizaje automático para la predicción de la presión de cierre de la fractura, closureTime, permeabilidad y tiempo hasta regímenes de flujo tardío de DFIT*. Unconventional resources technology conference.
- Carrera Jácome, L. (2018) *Análisis Nodal*. Obtenido de <https://es.scribd.com/document/393404481/Analisis-Nodal>
- Jami, O. (2019) *Alternativas para optimizar el sistema Power Oil en la estación Shushuqui*. Quito.
- MetaQuotes (04 de Noviembre de 2019) *MetaQuotes*. <https://www.mql5.com/es/articles/497>
- Ministerio de Energía y Recursos Naturales no Renovables. (2019) *Ministerio de Energía y Recursos Naturales no Renovables*. Obtenido de Ministerio de Energía y Recursos Naturales no Renovables: <https://www.recursoyenergia.gob.ec/>
- NumPy Community (05 de Noviembre de 2019) *What is NumPy*. <https://numpy.org/doc/1.17/user/whatisnumpy.html>.
- PyData Development Team (05 de Noviembre de 2019) *Package overview*. <https://pandas.pydata.org/pandas-docs/stable/>.
- Rahmanifard, H., & Plaksina, T. (2018) *Application of artificial intelligence techniques in the petroleum industry: a review*. USA.
- Scikit-learn (09 de Noviembre de 2019) *Scikit-Learn Machine Learning in Python*. Obtenido de Scikit-Learn Machine Learning in Python: <https://scikit-learn.org/stable/>.
- TensorFlow (2019 de Noviembre de 2019) *Why TensorFlow*. <https://www.tensorflow.org/about>.
- Zhou, P. (2019) *Production Data Analysis By Machine Learning*.