Humanidades, Tecnología y Ciencia, del Instituto Politécnico Nacional
Certificado de Reserva de Derechos al Uso Exclusivo del Título: 04-2010-0326124-14000-203 ISSN: 2007-1957

PREDICCIÓN DEL NÚMERO DE OCTANO (RON) EN GASOLINAS MEDIANTE APRENDIZAJE MAQUINAL Y REDES NEURONALES

Miriam Noemi Moreno Montiel
Instituto Politécnico Nacional - ESIQIE
mimorenom@ipn.mx

Angel Gabriel Garcia Medrano
Instituto Politécnico Nacional - ESIQIE
agarciam1821@alumno.ipn.mx

Eder Oscar Fajardo Valencia
Instituto Politécnico Nacional - ESIQIE
Efajardov1800@alumno.ipn.mx

Abstract

La determinación del índice antidetonante (AKI) es clave para la calidad de la gasolina, pero los métodos industriales actuales son costosos y lentos. Esta investigación desarrolló y validó un modelo de ***red neuronal artificial (ANN)*** *para predecir el Número de Octano de Investigación (RON) de forma rápida y eficiente.*

Utilizando 703 muestras, el modelo emplea como predictoras propiedades fisicoquímicas asequibles: densidad, azufre y puntos de destilación (T10, T50, T90). La red neuronal profunda diseñada demostró un rendimiento robusto, alcanzando un ***Error Absoluto Medio (MAE) de 2.77*** *en el conjunto de prueba. Estos hallazgos validan el enfoque computacional como una alternativa eficaz para optimizar el control de calidad.*

Palabras clave: Número de Octano, Machine Learning, Redes Neuronales, Gasolina, Predicción.

La gasolina, una compleja mezcla de hidrocarburos obtenida del petróleo, debe cumplir con estrictas especificaciones para su comercialización y para garantizar el

rendimiento óptimo de los motores de combustión interna las cuales son impuestas generalmente por la NOM-016-CREE-2016. Uno de los factores que se cuantifican para clasificar la calidad de la gasolina es el índice AKI (Anti-Knock Index) el cual teóricamente se sabe que es una variable que depende de los dos índices más, el MON (Motor Octane Number) y el RON (Research Octane Number), el índice AKI cumple la siguiente formula:

$$AKI = \frac{MON + RON}{2}$$

La calidad de la gasolina se clasifica comúnmente según este índice, como Premium (superior a 91) y Regular (entre 85 y 91).

El método de referencia para medir el RON es la norma ASTM D2699, que utiliza un motor especializado conocido como CFR (*Cooperative Fuel Research*) para carga ligera y el MON se mide con el método ASTM D2700 para condiciones de carga pesada. A pesar de la fiabilidad y estandarización de estos métodos, uno de los problemas actuales son los altos costos de operación, mantenimiento y de igual manera los altos tiempos de cuantificación para cada prueba, que como bien se sabe, en campo el factor del tiempo es esencial preservar, anteriormente no era posible innovar en este tema debido a la escases de tecnología con la que se contaba, pero actualmente se cuenta con una cantidad de datos histórica obtenida a partir de cientos de miles de pruebas llevadas a cabo, esto permite formular un método que optimice el tiempo de resultados para cada muestra reduciendo significativamente los costos de operación y/o mantenimiento, al mismo tiempo se da un aporte a la industria 4.0 en refinerías, terminales de almacenamiento y distribución, centros de despacho o gasolineras a nivel nacional.

La estrategia de esta investigación se centra en superar limitaciones mediante el desarrollo de un método computacional basado en *Machine Learning*. Aprovechando la disponibilidad de datos históricos de laboratorio, se propone el diseño de una red neuronal artificial (ANN) para predecir el AKI a partir de otras propiedades fisicoquímicas que son más asequibles y rápidas de obtener como son: densidad, contenido de azufre y los cortes de la curva de destilación T10, T50 y T90. La importancia de esta investigación es la optimización de costos y tiempos de operación sin sacrificar la fiabilidad de los resultados, ofreciendo una herramienta computacional que facilite la toma de decisiones y mejore el control de calidad en la producción de gasolinas.

El marco conceptual de este trabajo se basa en buscar una relación entre las propiedades fisicoquímicas de la gasolina y el AKI. El AKI está intrínsecamente ligado a la estructura de los hidrocarburos que componen la gasolina (parafinas, isoparafinas, aromáticos, etc.). Las propiedades seleccionadas cuidadosamente como predictoras actúan como indicadores indirectos (*proxies*) de esta composición. La densidad, por ejemplo, se correlaciona con la concentración de compuestos aromáticos, que tienen un alto RON. La curva de destilación (T10, T50, T90) caracteriza la volatilidad y la distribución de pesos moleculares de la

mezcla, factores que influyen en el comportamiento de la combustión.

Finalmente, el contenido de azufre, aunque no afecta directamente al RON, es un indicador del tipo de procesamiento de desulfuración que ha sufrido la gasolina, procesos que pueden alterar la composición de hidrocarburos y, por ende, el octanaje. La relación entre estas variables y el RON es compleja y no lineal, lo que justifica el uso de un modelo avanzado como una red neuronal, capaz de aprender estos patrones intrincados.

El propósito de esta investigación es, por lo tanto, desarrollar y validar dicho modelo predictvo.

1. Metodología

El estudio se llevó a cabo siguiendo un flujo de trabajo sistemático en literatura que abarca el preprocesamiento de los datos, el diseño y entrenamiento del modelo, y su posterior evaluación.

1.1. Conjunto de Datos

La presente investigación se fundamenta en una base de datos histórica que contiene mediciones de laboratorio de 703 muestras de gasolina, sin celdas en blanco. Para cada muestra, se registraron cinco variables predictoras y una variable objetivo:

- **Variables Predictoras (Entrada):**
  - Densidad (g/cm³)
  - Contenido de Azufre (ppm)
  - Punto de destilación T10 (°C)
  - Punto de destilación T50 (°C)
  - Punto de destilación T90 (°C)

- **Variable Objetivo (Salida):**
  - Número de Octano de Investigación (RON)

Tabla 1. Estadísticas descriptivas del conjunto de datos inicial.

	Octane raiting	Density (g/cm³)	Sulfur (ppm)	T10 (°C)	T50 (°C)	T90 (°C)
count	703.000000	703.000000	703.000000	703.000000	703.000000	703.000000
mean	90.443528	0.733589	42.251238	54.189047	95.542674	155.029445
std	4.652731	0.020444	264.446696	7.186120	10.777736	13.359356
min	72.900000	0.673800	3.260000	40.100000	80.000000	130.000000
25%	88.000000	0.721100	15.495000	49.300000	86.650000	144.350000
50%	89.600000	0.731200	24.950000	53.700000	92.900000	155.300000
75%	93.000000	0.744250	39.100000	58.750000	103.450000	165.700000
max	106.000000	0.918600	6993.000000	69.800000	120.000000	180.000000

1.2. Procedimientos implementados:

Preprocesamiento de Datos: Antes del entrenamiento, los datos fueron normalizados para asegurar que todas las características contribuyeran de manera equitativa al aprendizaje del modelo. Las redes neuronales son sensibles a la escala de los datos de entrada, y este paso es crucial para una convergencia eficiente. Se aplicaron dos técnicas de escalado:

StandardScaler: Para las variables de entrada (X), a cada característica se le restó su media y se dividió por su desviación estándar.

Normalización Min-Max: La variable objetivo (Y, RON) se escaló a un rango entre 0 y 1 utilizando la fórmula yscaled=(y−ymin)/(ymax−ymin esto debido a que las redes

neuronales funcionan mejor cuando hay un rango definido que no sea tan grande ni pequeño.

2. Diseño del Modelo de Red Neuronal:

Se implementó una Red Neuronal Profunda (ANN) utilizando la librería Keras de TensorFlow en Python versión 3.11. La arquitectura del modelo se diseñó para capturar las relaciones no lineales complejas presentes en los datos.

Arquitectura: Se construyó un modelo secuencial con tres capas ocultas densas en una configuración decreciente (256 → 128 → 64 neuronas).

Función de Activación: Se utilizó la función Unidad Lineal Rectificada (ReLU) en todas las capas ocultas, una elección estándar en redes profundas por su eficiencia computacional y su capacidad para mitigar el problema del gradiente desvaneciente.

Capa de Salida: La red finaliza con una única neurona de salida con activación lineal, adecuada para un problema de regresión de valor único.

3. Compilación y Entrenamiento:

Se configuró el proceso de aprendizaje del modelo mediante:
Optimizador: Se eligió el optimizador Adam con una tasa de aprendizaje de 0.0001, conocido por su capacidad de ajustar la tasa de aprendizaje de forma adaptativa para cada parámetro del modelo.

Función de Pérdida: El modelo fue entrenado para minimizar el Error Cuadrático Medio (MSE, *Mean Squared Error*), una métrica estándar que nos permite evaluar la confiabilidad del modelo propuesto.

Control del Entrenamiento: Para prevenir el sobreajuste y asegurar que el modelo generalice bien a datos no vistos, se implementó un mecanismo de parada temprana. Este monitorea la pérdida en el conjunto de validación (val_loss) y detiene el entrenamiento si no se observa una mejora después de 50 épocas consecutivas, restaurando los pesos del modelo correspondientes al punto de mejor rendimiento.

4. Evaluación del Modelo:

La capacidad predictiva y la confiabilidad del modelo se evaluaron mediante dos estrategias:

Validación Cruzada: Se aplicó una validación cruzada de 5 pliegues (*5-fold cross-validation*) sobre el conjunto de datos completo de 703 muestras. Este método proporciona una estimación robusta del rendimiento del modelo en datos no vistos.

Prueba Final: Adicionalmente, se realizó una división única del conjunto de datos en 80% para entrenamiento y 20% para prueba. El modelo se entrenó con el 80% y su rendimiento final se midió sobre el 20% de datos que nunca vio durante el entrenamiento.

Métrica de Evaluación: La métrica principal para evaluar el rendimiento fue el Error Absoluto Medio (MAE), que se convirtió de la escala normalizada (0-1) a la escala original del RON para una interpretación directa.

5. Resultados:

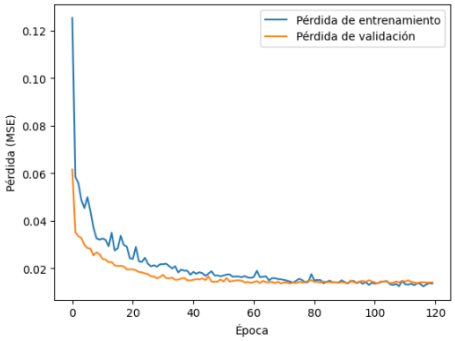
El análisis exploratorio inicial reveló las relaciones lineales entre las variables. La densidad y el T50 mostraron las correlaciones positivas más fuertes con el RON, con coeficientes de 0.38 y 0.32, respectivamente. El contenido de azufre presentó una correlación casi nula (-0.01). Esta baja correlación lineal general subraya la necesidad de un modelo no lineal para capturar la complejidad de las interacciones.

Tabla 2. Resultados de análisis exploratorio en Python 3.11.

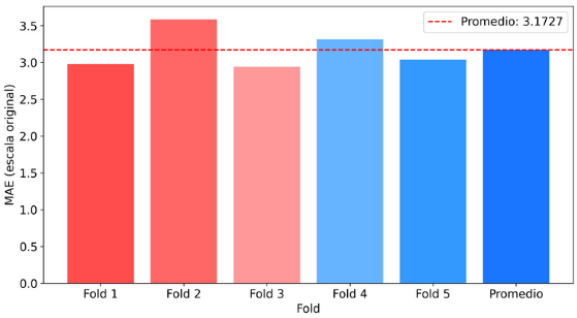
Octane raiting	1.000000
Density (g/cm³)	0.381222
Sulfur (ppm)	-0.013451
T10 (°C)	0.125261
T50 (°C)	0.323958
T90 (°C)	0.094790

El entrenamiento del modelo de red neuronal mostró una convergencia, como se observa en la gráfica 1. Las curvas de pérdida de entrenamiento y validación disminuyen y se estabilizan juntas, lo que indica que el modelo aprendió los patrones de los datos de manera efectiva sin sobreajustar significativamente al conjunto de entrenamiento.

Grafica 1. Perdidas de entrenamiento y validación por época.



La evaluación mediante validación cruzada de 5 pliegues confirmó la consistencia y robustez del modelo. Como se muestra en la Figura 2, el Error Absoluto Medio (MAE) en la escala original del RON fue estable a través de los diferentes pliegues, con valores que oscilaron entre 2.94 y 3.59. El MAE promedio de la validación cruzada fue de **3.17**, lo que demuestra que el rendimiento del modelo no depende de una partición específica de los datos y es generalizable.

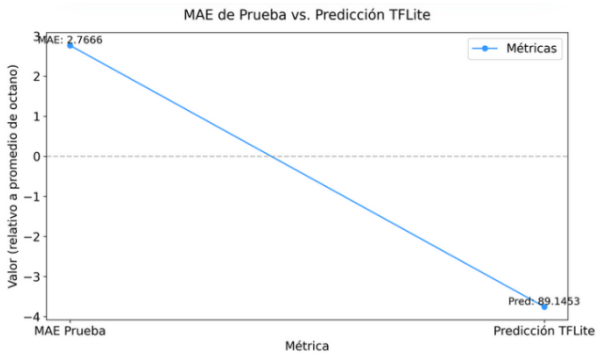


Grafica 2. Error absoluto medio (MSE) en validación cruzada por "Fold".

En la prueba final, utilizando el 20% de los datos reservados, el modelo alcanzó un **MAE de 2.77** en la escala original del RON. Este resultado se considera un desempeño

razonable y valida la capacidad predictiva del modelo en datos completamente nuevos.

visualmente en la gráfica 3, que compara los valores de RON predichos por el modelo con los valores reales medidos en el laboratorio. Los puntos se agrupan de forma compacta a lo largo de la línea diagonal, que representa una predicción perfecta. Esta fuerte alineación indica que el modelo no presenta sesgos evidentes y funciona bien en todo el rango de valores de octanaje analizado.



Grafica 3. Primera prueba de red neuronal.

6. Discusión:

Los resultados presentados demuestran que el modelo de red neuronal profunda es capaz de predecir el Número de Octano de Investigación (RON) con un grado de precisión razonable, validando la hipótesis central de esta investigación. El MAE final de 2.77 unidades de RON es un indicador prometedor de la viabilidad de este enfoque como herramienta de apoyo en el control de calidad de las refinerías. Al contextualizar este resultado con la literatura, se observa que, si bien es un rendimiento sólido, existe margen para futuras mejoras. Por ejemplo, estudios que utilizan arquitecturas híbridas de ANN con algoritmos genéticos han reportado un MAE de 1.4 para el índice antidetonante (AKI) , y otros trabajos comparativos han encontrado que modelos como las Máquinas de Vectores de Soporte (SVR) pueden alcanzar un MAE tan bajo como 0.9.[17] Esto sugiere que la exploración de algoritmos alternativos como XGBoost o SVR podría ser una vía fructífera para refinar aún más la precisión predictiva.

Una de las consecuencias teóricas más importantes de este estudio es la confirmación empírica de que un modelo de ANN puede capturar eficazmente las complejas y no lineales interacciones entre las propiedades fisicoquímicas de la gasolina y su RON. El análisis de correlación inicial mostró que las relaciones lineales eran débiles, lo que hacía prever el fracaso de modelos más simples. El éxito de la ANN radica en su capacidad para aprender representaciones jerárquicas de los datos a través de sus capas ocultas, modelando sinergias y antagonismos entre la densidad, el azufre y la volatilidad que determinan el comportamiento final de la combustión. La validez de las conclusiones se ve reforzada por la consistencia de los resultados en la validación cruzada, que indica que el modelo es robusto y su rendimiento es generalizable a nuevas muestras de gasolina con características similares.

Las implicaciones prácticas de esta investigación son significativas. La implementación de este modelo en un entorno de producción podría transformar el control de calidad. Permitiría a los ingenieros de proceso obtener una estimación casi instantánea del RON, en lugar de esperar horas por el resultado del motor CFR. Esto no solo reduce

costos operativos, sino que también permite un ajuste más rápido y preciso de las mezclas de gasolina, minimizando el "regalo" de octanaje (*octane giveaway*) y asegurando que el producto final cumpla consistentemente con las especificaciones. A largo plazo, la integración de este tipo de herramientas predictivas es un paso hacia la digitalización y optimización de los procesos de refinación.

7. Siguientes pasos por seguir.

Los resultados obtenidos en esta investigación sientan las bases para avanzar hacia fases más aplicadas y orientadas a la integración tecnológica. Los próximos pasos propuestos son los siguientes:

Desarrollo de una aplicación universal

Implementar una herramienta computacional accesible:
(App/WebApp) "Pedrict Ron" que permita al usuario ingresar de manera directa las variables de entrada (densidad, azufre, T10, T50, T90) y obtener un valor fiable de AKI de cualquier muestra. Desarrollar la interfaz utilizando **JavaScript** para el entorno de usuario y **SQL** para la gestión de la base de datos, con esto asegurar compatibilidad multiplataforma para facilitar su uso en refinerías, laboratorios y estaciones de servicio.

Alimentación dinámica de la base de datos

Establecer un sistema que permita actualizar la base de datos en tiempo real con nuevas muestras experimentales.
Garantizar un esquema de **aprendizaje continuo** del modelo, fortaleciendo su capacidad predictiva con cada actualización.

Validación en entornos reales de producción

Probar la herramienta en plantas de refinación y terminales de almacenamiento para evaluar su desempeño bajo condiciones operativas reales.

Realizar pruebas de campo comparando las predicciones de la aplicación con los resultados obtenidos mediante el método CFR (ASTM D2699 y ASTM D2700).

Escalabilidad e integración en Industria 4.0

Incorporar la aplicación en los sistemas de control de calidad y gestión de procesos de refinerías.

Explorar la integración con tecnologías de **IoT (Internet of Things)** y **sistemas SCADA**, orientados a la automatización y optimización de la producción.

Referencias

Alboqami, F., van Oudenhoven, V. C. O., Sarathy, S. M., Tzanetakis, T., and Kontoravdi, C. (2022). A case study on integrating an AI system into the fuel blending process in a chemical refinery Processes 9 4

Faraggi, D., and Simon, R. (1995). A neural network model for survival data Stat. Med. 14 73 82

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine Ann. Stat. 29 1189 1232

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutsever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. Recuperado el 19 de agosto de 2025, de https://arxiv.org/abs/1207.0580

Kessler, T. J., Hudson, C., Whitmore, L., and Mack, J. H. (2022). Prediction of Research/Motor Octane Number and Octane Sensitivity Using Artificial Neural Networks. Recuperado el 19 de agosto de 2025, de https://doi.org/10.31224/osf.io/59hfp

Li, Z., Yuan, Z., Li, L., and Zhang, Y. (2021). Comparison of machine learning algorithms for predicting the research octane number of gasoline blends Energy Fuels 35 9497 9506

Livingstone, D. J. (2008). Artificial neural networks. Humana Press.

Ryan, T. W., and Callahan, T. J. (1992). Engine and piston design and their effects on lubrication and fuel economy. Recuperado el 19 de agosto de 2025, de https://doi.org/10.4271/920488

Srivastava, N., Hinton, G., Krizhevsky, A., Sutsever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting J. Mach. Learn. Res. 15 1929 1958

Sun, X., Zhang, F., Liu, J., and Xiongbo, D. (2022). Prediction of gasoline research octane number using multiple feature machine learning models Fuel 324 124567

U.S. Energy Information Administration. (s.f.). Octane in-depth. Recuperado el 19 de agosto de 2025, de https://www.eia.gov/energyexplained/gasoline/octane-in-depth.php

Vapnik, V. (1995). The nature of statistical learning theory. Springer-Verlag.

Wikipedia contributors. (2024). Octane rating. Recuperado el 19 de agosto de 2025, de https://en.wikipedia.org/wiki/Octane_rating

Yuan, Z., Li, L., Zhang, Y., and Li, Z. (2022). A prediction model of RON loss based on neural network. Recuperado el 19 de agosto de 2025, de https://doi.org/10.4271/2022-01-0418

Zhang, Y., Li, L., Yuan, Z., and Li, Z. (2022). Prediction of research octane number loss and sulfur content in gasoline refining using machine learning Fuel 310 122456

Zhao, L., Zhang, Y., Li, L., and Yuan, Z. (2022). A hybrid intelligent method for predicting gasoline octane number and optimising operation parameters Chem. Eng. Trans. 94 1165 1170