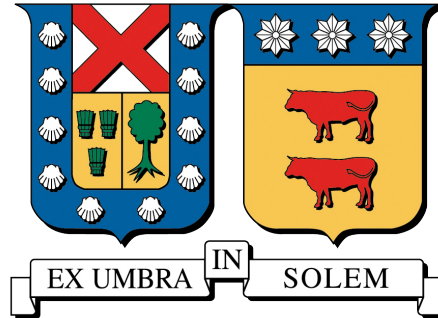


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS
SANTIAGO - CHILE



**DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINACIÓN DE
CAPACIDAD ANTIDETONANTE EN COMBUSTIBLES DE AVIACIÓN**

TESIS DE GRADO PRESENTADA POR

FELIPE ANDRÉS VALDIVIA CÁCERES

COMO REQUISITO PARCIAL PARA OPTAR AL GRADO DE
MAGISTER EN INNOVACIÓN Y EMPRENDIMIENTO

PROFESOR GUÍA : SR. CLAUDIO GUILLERMO RAMIREZ MORA
PROFESOR CORREFERENTE : SR. ESTEBAN SIMÓN PROGASKA SILVA

SEPTIEMBRE 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título; Tesis de Postgrado;

Título del trabajo: DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINACIÓN DE
CAPACIDAD ANTI-DETONANTE EN COMBUSTIBLES DE AVIACIÓN

Nombre del candidato(a): Felipe Andrés Valdivia Cáceres

Carrera / Grado: Magíster en Innovación Tecnológica y Emprendimiento

Campus: Santiago Vitacura ; **Departamento:** Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Claudio Ramirez Mora, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

6 meses; 12 meses; 2 años; 3 años; 5 años; 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 06-10-2025 ; Firma: 

Estudiante o Candidato(a):

Fecha: 06-10-2025 ; Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

A mi familia ...

Por darme su tiempo para realizar este estudio.

RESUMEN EJECUTIVO

En aviación existen dos tipos de impulsión: motores de pistón o ignición por chispa y motores de turbina. La Gasolina de Aviación (AvGas) es un combustible para la aviación basada en motores de pistón como las avionetas. La producción de esta gasolina es acotada debido a la demanda mensual. La calidad y seguridad de la gasolina de aviación dependen de varios parámetros o análisis que se hacen en el combustible y uno de ellos es la capacidad antidetonante o Detonancia de mezcla rica, medida tradicionalmente mediante el método ASTM D909-22 con un motor CFR Waukesha, un equipo fabricado en 1959 que hoy presenta múltiples desafíos: alta complejidad, sin recambio y un proceso que requiere cerca de seis horas para entregar resultados. En Chile, solo un reducido número de especialistas domina esta técnica, lo que representa un riesgo operativo y regulatorio para la venta de la Gasolina de Aviación, producida exclusivamente en Refinería Aconcagua.

Esta tesina propone el desarrollo de un modelo predictivo basado en machine learning (ML) que permita estimar la Detonancia mezcla rica a partir de otras propiedades fisicoquímicas registradas en sistemas de laboratorio (LIMS). El modelo Random Forest resultó ser la alternativa más precisa en el conjunto de datos estudiado, con un error promedio inferior al 1 %, reduciendo drásticamente los tiempos de análisis desde horas a segundos y mejorando la eficiencia en el proceso de control de calidad de la Gasolina de Aviación.

Más allá del aspecto técnico, con este desarrollo se plantea un modelo de negocio escalable bajo la modalidad Machine Learning-as-a-Service (MLaaS), que convierte este piloto en la base para ofrecer servicios de analítica predictiva a laboratorios, refinerías y organismos reguladores. La propuesta se sustenta en ingresos recurrentes por suscripción y proyectos específicos de implementación, con potencial de expansión hacia la industria energética latinoamericana y otros sectores industriales.

En síntesis, este trabajo integra ciencia de datos, innovación y emprendimiento para transformar un procedimiento analítico crítico en una solución digital más rápida, económica y escalable, que refuerza la competitividad y sostenibilidad de la industria de combustibles en Chile y la región.

Palabras Clave. Método, Análisis, Predicción, Laboratorio, Control de Calidad, Machine learning, Industria energética, Innovación tecnológica

ABSTRACT

In aviation, there are two types of propulsion: piston or spark-ignition engines, and turbine engines. Aviation Gasoline (AvGas) is a fuel for aviation applications based on piston engines such as those used in light aircraft. Production of this gasoline is limited by monthly demand. The quality and safety of aviation gasoline depend on several parameters or analyses performed on the fuel, one of which is the anti-knock capacity or Rich Burn Knock, traditionally measured using the ASTM D909-22 method using a CFR Waukesha engine, a machine manufactured in 1959 that today presents multiple challenges: high complexity, no replacement parts, and a process that requires approximately six hours to deliver results. In Chile, only a small number of specialists master this technique, which represents an operational and regulatory risk for the sale of Aviation Gasoline, produced exclusively at the Aconcagua Refinery.

This thesis proposes the development of a machine learning (ML)-based predictive model that allows estimating Detonation from other physicochemical properties recorded in laboratory-based systems (LIMS). The Random Forest model proved to be the most accurate alternative in the studied dataset, with an average error of less than 1 %. This model drastically reduced analysis times from hours to seconds and improved efficiency in the Aviation Gasoline quality control process.

Beyond the technical aspect, this development proposes a scalable business model under the Machine Learning-as-a-Service (MLaaS) modality, making this pilot the basis for offering predictive analytics services to laboratories, refineries, and regulatory agencies. The proposal is supported by recurring subscription revenue and specific implementation projects, with potential for expansion into the Latin American energy industry and other industrial sectors.

In summary, this work integrates data science, innovation, and entrepreneurship to transform a critical analytical procedure into a faster, more economical, and scalable digital solution, strengthening the competitiveness and sustainability of the fuel industry in Chile and the region.

Keywords. Method, Analysis, Prediction, Laboratory, Quality Control, Machine Learning, Energy industry, Technological innovation

Índice de Contenidos

1. Introducción	1
2. Objetivos	4
2.1. Objetivos Generales	4
2.2. Objetivos específicos	4
3. Alcance	5
4. Metodología y Desarrollo	6
4.1. Objeto de estudio	6
4.2. Estado del Arte	6
4.2.1. Aviación Civil en Chile	6
4.2.2. Producción de Combustible de Aviación	7
4.2.3. Consumo de Combustible de Aviación	8
4.2.4. El método estándar D909-22	9
4.2.5. Enfoques predictivos en la industria	12
4.3. Metodología utilizada	13
4.4. Herramientas y entornos	14
4.5. Limitaciones	14
5. Desarrollo Metodológico	15
5.1. Análisis basado en ML para la Detonancia Mezcla Rica en Gasolina de Aviación	15
5.1.1. Recolección de datos	15
5.1.2. Preparación de los datos	15
5.1.3. Selección de variables	16
5.1.4. Desarrollo de modelos y evaluación	18
5.1.5. Propuesta de implementación	18
5.2. Modelo de Negocio	19
5.2.1. Análisis basado en ML como servicio	19
5.2.2. Breve mirada al Mercado	20
5.2.3. Canvas	21
5.2.3.1. Segmentos de Cliente	21
5.2.3.2. Propuesta de Valor	21
5.2.3.3. Canales de Distribución y Comunicación	22
5.2.3.4. Relación con el Cliente	23
5.2.3.5. Fuentes de Ingreso	23
5.2.3.6. Recursos Clave	23
5.2.3.7. Actividades Clave	23
5.2.3.8. Socios Clave	23
5.2.3.9. Estructura de Costos	24
5.2.4. Flujo de caja	24

6. Resultados Obtenidos	26
6.1. Modelo de ML	26
6.2. Modelo de Negocio	29
7. Conclusiones	32
7.1. Declaraciones Generales	32
7.2. Reflexión Final	32
Bibliografía	33
A. Especificación de Calidad de Gasolina de Aviación	34
B. Código Fuente Notebook Jupyter	36

Índice de Tablas

4.1. Valores por región y año	10
4.2. Total de ventas Gasolina de Aviación por mes y año	10
5.1. Total de Ingresos (Millones de pesos)	25
5.2. Total de Egresos y Diferencia de caja (Millones de pesos)	25
6.1. Comparación de desempeño de los modelos en entrenamiento (70 % de los datos) y test (30 % de los datos).	26
6.2. Comparación del desempeño del modelo Random Forest en entrenamiento y test	28
6.3. Comparación entre valores reales de detonancia mezcla rica y predicciones del modelo.	28
6.4. Balance General Realista (en Millones de pesos)	30
6.5. Tasas y situaciones proyectadas bajo escenarios Real, Pesimista y Optimista	30
6.6. Escenarios de evaluación: Real, Pesimista y Optimista (Millones de pesos)	31
6.7. Valoración de la empresa en distintos escenarios (Millones de pesos)	31

Índice de Figuras

1.1. F4 Análisis de Supercarga, Waukesha	2
4.1. Piper J5	7
4.2. Evolución diaria de una campaña de Gasolina de Aviación	8
4.3. Certificado de Análisis ejemplo	9
4.4. Total de Ventas de Gasolina de Aviación 2023 al 2025	11
4.5. Desarrollo de curvas de potencia limitada por detonación	13
5.1. Resultados de muestra en SampleManager, LIMS	16
5.2. Mapa de Calor Resultados de variables analizadas	17
5.3. Modelo de Implementación Técnica	19
5.4. Hype Cycle Inteligencia Artificial 2025	20
5.5. Mercado Global de Machine Learning 2025 al 2034	21
5.6. Mercado por región de Machine Learning 2024	22
5.7. Canvas Bussines Model	22
A.1. Requisitos y especificaciones nacionales de calidad para gasolina de aviación	35

1 | Introducción

En los aviones existen dos tipos de impulsión: motores de pistón y motores de turbina. La Gasolina de Aviación (AvGas) es un combustible para la aviación basada en motores de pistón (o también llamados motores de ignición por chispa) como las avionetas, en contraparte del Kerosene de Aviación (Kero Jet) que utilizan los motores de turbina. La producción de esta gasolina se realiza principalmente en la Refinería Aconcagua de la Empresa Nacional del Petróleo, ENAP. La distribución de este combustible a todo Chile depende de empresas como Copec y Esmax.

Dentro de las propiedades físico-químicas que se necesitan para asegurar los requisitos de calidad y seguridad para la venta del producto se encuentra la Detonancia Mezcla Rica. Detonancia se refiere al golpeteo del motor o "knock" que es cuando la mezcla de aire y combustible en el cilindro explota de golpe y fuera de control, en lugar de quemarse de manera uniforme con la chispa. La Mezcla Rica se refiere a que hay más gasolina que aire del necesario. Una mezcla rica se usa a veces para evitar sobrecalentamiento y mejorar la potencia, pero consume más y produce más emisiones. Mientras más resistente sea la gasolina a detonar, mejor es su calidad (más octanos).

Para determinar la propiedad de la Detonancia Mezcla Rica se utiliza un procedimiento basado en el método ASTM¹ D909-22 ([ASTM Internacional \(2022\)](#)) que consiste en utilizar un motor de prueba en donde se hace correr la gasolina, se le va cambiando la mezcla aire-combustible y se mide cuánta potencia soporta antes de detonar. Luego se compara con combustibles conocidos y así se le asigna su "puntaje" de calidad y por esto se denomina en la normativa como desempeño supercarga (según procedimiento interno [ENAP \(2023\)](#)).

El motor donde se realiza la prueba tiene varias denominaciones, pero la más común es referirse a este como Motor CFR²Cooperative Fuel Research) Waukesha³ (Figura 1.1) y es un motor de un cilindro, ciclo de 4 tiempos, cabezal ajustable (lo que permite variar la relación de compresión fácilmente) y está refrigerado por agua y diseñado solo para pruebas de combustibles, no para uso comercial.

Es en este motor donde se produce el problema: un análisis completo demora más de 6 horas en ejecutar, es una máquina antigua y no tiene recambio debido a que los nuevos motores que venden en el mercado ya no contemplan la implementación del ASTM D909-22 y por lo tanto, en caso de tener una indisponibilidad del motor actual, *no se podría certificar ni vender la Gasolina de Aviación*, provocando un problema en la cadena de suministro al parque de aviación de motores de pistón.

Cumplir con la norma ASTM D909-22 para esta propiedad de Detonancia mezcla rica es esencial para garantizar que los combustibles utilizados en la aviación de pistón cumplan con los requisitos de calidad y seguridad, que a su vez es controlado por el ente regulador, Superintendencia de Electricidad y Combustibles (SEC). La capacidad del combustible para evitar la detonación es vital en estos motores de alta compresión, lo que hace que el número de octano medido bajo este método sea un indicador clave del rendimiento y la seguridad operativa.

La idea principal de esta tesis es obtener una estimación por medio de otros resultados analíticos del combustible, que nos permita en el largo plazo desplazar o tener una alternativa al método D909-22.

¹American Society for Testing and Materials

²(

³Marca del motor



Figura 1.1: F4 Análisis de Supercarga, Waukesha

(Fuente: Laboratorio ENAP Aconcagua)

Actualmente, los laboratorios de calidad cuentan con una cantidad de información histórica registrada en sistemas como LIMS⁴, que incluyen múltiples propiedades físico-químicas de las muestras de gasolina de aviación analizadas a lo largo del tiempo. Este contexto habilita una oportunidad para aplicar herramientas de análisis de datos y machine learning que permitan modelar y predecir el comportamiento antidetonante de dichas mezclas, basándose en variables medibles de forma rutinaria.

No obstante, este caso específico de Machine learning, por su carácter especializado y de nicho, no es por sí mismo escalable como producto comercial masivo. Su verdadero valor radica en constituir un caso piloto, demostrativo de un enfoque metodológico que puede ser replicado, paquetizado y extendido hacia otros análisis de laboratorio y, en una fase más avanzada, hacia la predicción de propiedades en tiempo real en planta mediante la integración con sistemas de control y monitoreo (PI System, DCS, SCADA).

Este trabajo, por tanto, persigue un doble propósito:

- Técnico, mediante el diseño, entrenamiento y validación de un modelo de machine learning capaz de predecir la Detonancia Mezcla Rica con precisión aceptable.
- Estratégico, mediante la formulación de un modelo de negocio escalable, que permita transformar este enfoque en una solución paquetizada, replicable y vendible bajo esquemas de suscripción o

⁴Laboratory Information Management System

licenciamiento, con potencial de adopción en la industria energética nacional e internacional.

De este modo, la presente tesina, que se enmarca dentro del Magíster en Innovación Tecnológica y Emprendimiento (MITE) de la Universidad Técnica Federico Santa María, no solo aborda un desafío tecnológico concreto como piloto, sino que también plantea una visión de innovación aplicada basada en datos, donde la digitalización, el análisis de datos y el emprendimiento convergen para ofrecer soluciones sostenibles y de alto impacto en sectores regulados y de alta criticidad como la energía, la aviación y otros.

2 | Objetivos

2.1. Objetivos Generales

Desarrollar un modelo piloto basado en Machine Learning que permita una predicción de la Detonancia de mezcla rica en base a otras propiedades de la Gasolina de Aviación y formular un modelo de negocio a partir de esta metodología.

2.2. Objetivos específicos

- Desarrollar el modelo que permita identificar propiedades relevantes para el cálculo.
- Implementar el algoritmo para su utilización en el Laboratorio.
- Diseñar un modelo de negocio que permita proveer este desarrollo como servicio.

3 | Alcance

El desarrollo de la tesina queda circunscrito a la elaboración de un modelo de Machine Learning piloto para la obtención del análisis de *Detonancia mezcla rica* que normalmente se obtiene por el método analítico ASTM D909-22 utilizando un equipo de laboratorio llamado Motor CFR Waukesha. Con este desarrollo piloto se diseña un modelo de negocio escalable para la venta de servicios analíticos basados en la paquetización, metodología y aprendizaje utilizado durante el desarrollo del modelo de machine learning.

4 | Metodología y Desarrollo

4.1. Objeto de estudio

El objeto de estudio corresponde al conjunto de muestras históricas de gasolina de aviación analizadas en un laboratorio de calidad en un recinto industrial, cuyas propiedades fisico-químicas están registradas en el sistema de gestión de información de laboratorio (LIMS). Estas muestras corresponden a mezclas reales producidas en refinería, utilizadas en la aviación de motor de pistón civil o militar, y sujetas a estándares normativos de ASTM.

El objetivo es predecir el comportamiento antidetonante (valor comparable al entregado por el método ASTM D909-22 con motor CFR Waukesha) a partir de un conjunto de variables disponibles en los registros analíticos, tales como presión de vapor, número de octano motor (MON), densidad, curva destilación, plomo, entre otros.

A partir de este caso específico, se busca generar una metodología replicable y escalable, que se puede aplicar a otros análisis de laboratorio y eventualmente a sistemas en tiempo real de planta.

4.2. Estado del Arte

4.2.1. Aviación Civil en Chile

Dentro del marco de la Aeronáutica Civil y Militar existen dos grupos de aviones o avionetas, las de motor a pistón (ignición por chispa) y las de motor de turbina. Para cada caso se utiliza un combustible específico. Para el primero, Gasolina de Aviación (AvGas) y para el segundo, Kerosene de Aviación (Jet Fuel).

Las aeronaves con motor de pistón son generalmente pequeñas y son esenciales para tareas y misiones que pueden comprender desde el entrenamiento de pilotos hasta traslado de enfermos o uso en Agricultura. Este tipo de nave están presentes en todo Chile y de hecho tienen un *componente social*, que hace necesario su abastecimiento de combustible por medio de campañas de producción en la Refinería Aconcagua. Su posterior distribución la realiza principalmente Copec y Esmax, en aeropuertos y aeródromos.

Según datos de la [Dirección General de Aeronáutica Civil \(2019\)](#), DGAC, existen en el Registro Nacional de Aeronaves (RNA) un total de 1.841 aeronaves registradas, de las cuales 1.195 son de propiedad privada, 642 de uso comercial, 30 pertenecientes a Carabineros y 4 de la PDI. Uno de los primeros registros fue un Piper como el de la figura 4.1.

Por sus características, economía y movilidad⁵, estas aeronaves cumplen las siguientes funciones:

- Traslados particulares

⁵Necesitan infraestructura mínima para despegue/aterrizaje



Figura 4.1: Piper J5

(Fuente: Dirección General de Aeronáutica Civil)

- Traslado de enfermos y órganos
- Combate al fuego
- Vigilancia, fotografía aérea
- Agricultura, fumigación
- Emergencias
- Entrenamiento en vuelo
- Transporte de cargas

4.2.2. Producción de Combustible de Aviación

El combustible necesario para el motor de pistón es la Gasolina de Aviación o AvGas, normalmente de 100 Octanos con bajo contenido de plomo. Dichas características lo hacen un combustible especial y normalmente se produce en campañas de esta gasolina para llenar un estanque específico para tal fin en la Refinería Aconcagua. Cada campaña suele demorar entre 4 a 5 días para una producción de 400m³ de este producto, en donde se van ejecutando las mezclas para obtener el Octanaje y propiedades adecuadas.

Si se observa la Figura 4.2 se muestra un proceso de campaña, donde se va mezclando en un tanque determinado hasta la altura definida. Durante este proceso de producción o *Preparación*, a medida que se llena el tanque, se toman muestras y se realizan análisis de laboratorio que son críticos para el control del producto final⁶:

- D4176, Apariencia
- D4052, API/Densidad Digital

⁶Cada análisis ASTM se nombra como DXXXX, por ejemplo D909-22

- D86, Evaporado Aviation
- D5453, Azufre
- D5191, Presión de Vapor
- D130, Corrosión en Hidrocarburos
- D3341, Plomo
- Antioxidante
- Colorante Azul

Con estos análisis, se va controlando hasta llenar el tanque y cuando ocurre el tanque se certifica, requisito para entrega a ventas.

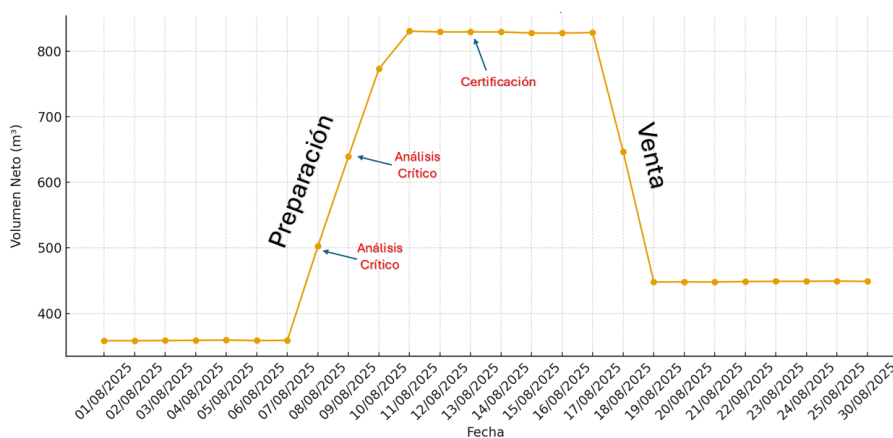


Figura 4.2: Evolución diaria de una campaña de Gasolina de Aviación

(Fuente: Elaboración propia)

En Chile el marco regulatorio lo dicta el Decreto 67 del [Ministerio de Energía \(2021\)](#) y es el encargado de controlar que se cumplan las especificaciones es la Superintendencia de Electricidad y Combustibles (SEC). Las propiedades que se certifican se pueden observar en la Figura 4.3 y para mayor referencia se incluye en el Anexo A la especificación de la Gasolina de Aviación del decreto. Si se compara con el listado anterior, notarán que ese listado es un subconjunto de esta especificación.

Para el cumplimiento de la producción y como etapa final del proceso, se emite un Certificado de Análisis o Informe de Ensayo, como el de la figura 4.3. Cada proceso de muestreo y análisis de los datos en Refinería, tanto crítico como de certificación final es documentado por el sistema LIMS que servirá de fuente de datos.

4.2.3. Consumo de Combustible de Aviación

No existen datos de consumo de Gasolina de Aviación, pero se puede inferir según las ventas. Cada venta en Chile, a minoristas y mayoristas, se debe registrar en un informe mensual a la SEC. Este informe contiene las cantidades vendidas de cada producto y sus calidades, de tal forma que la SEC puede verificar todos los combustibles que se transan en el mercado. En este aspecto y según estos registros de la SEC⁷, en los años 2022, 2023 y 2024 se venden en el país un promedio de 4.800 m³ anuales de Gasolina de Aviación 100-300 aproximadamente (recordamos los 400 m³ de producción mensual).

⁷<https://www.sec.cl/venta-mensual-combustibles-liquidos-y-gas-licuado-de-petroleo/>

INFORME DE ENSAYO				
Gasolina de Aviación 100 LL				
(Valores Referidos: DS 67/2020 Min. de Energía)				
NUMERO DE ENSAYO	:			
NUMERO DE MUESTRA	:			
TANQUE	:			
FECHA DE MUESTRA	:			
PROPIEDAD	UNIDAD	RESULTADO	REFERENCIA	METODO
API Gravedad, Digital 60°F	°API	71.5	Informar	ASTM D4052
Densidad, Digital 15°C	kg/m ³	696.9	Informar	ASTM D4052
Color	-	Pasa	Pasa	ASTM D2392
Corrosión Lámina de cobre (2 h a 100°C)	Nº	1b	Max: 1	ASTM D130
Reacción con agua/cambio	ml	0	Min: -2 Max: 2	ASTM D1094
Punto de Congelación	°C	<-73.0	Max: -58.0	ASTM D2386
Punto Inicial	°C	32.6	Informar	ASTM D86
10% Evaporado	°C	69.0	Max: 75	ASTM D86
40% Evaporado	°C	100.4	Min: 75	ASTM D86
50% Evaporado	°C	103.2	Max: 105	ASTM D86
90% Evaporado	°C	112.2	Max: 135	ASTM D86
Punto Final	°C	135.8	Max: 170	ASTM D86
10% + 50%	°C	172.2	Min: 135	ASTM D86
Recuperado	% v/v	97.7	Min: 97	ASTM D86
Residuo	% v/v	0.9	Max: 1.5	ASTM D86
Pérdida	% v/v	1.4	Max: 1.5	ASTM D86
Plomo Precipitado	mg/100ml	<1	Max: 3	ASTM D873
Goma Potencial	mg/100ml	<1	Max: 6	ASTM D873
Presión de vapor	psi	6.53	Min: 5.5 Max: 7.0	ASTM D5191
Calor Neto de Combustión	MJ/kg	44.406	Min: 43.500	ASTM D3338
Azufre	% m/m	<0.0005	Max: 0.05	ASTM D5453
Número de Octano Motor	-	106.2	Min: 99.60	ASTM D2700
Aviation Lean Rating	-	117.62	Min: 100.00	ASTM D2700
Detonancia mezcla rica	-	133.8	Min: 130	ASTM D909
Plomo	g Pb/l	0.48	Min: 0.28 Max: 0.56	ASTM D3341
Antioxidante	mg/l	2.357	Max: 12	-
Colorante Azul	mg/l	2.04	Max: 2.7	-

NOTA (1): Refinería Aconcagua no adiciona aditivos de conductividad eléctrica en la formulación de la Gasolina de Aviación.
 NOTA (2): Los resultados emitidos en este certificado se relacionan solo con la muestra indicada en él.
 NOTA (3): Fecha de análisis está comprendida entre la fecha de muestra y la fecha de emisión del certificado
 Muestreo realizado por terceros, de acuerdo con la Norma Chilena 60:1996 y procedimiento ERA-LAB-PRG-003.

Figura 4.3: Certificado de Análisis ejemplo

(Fuente: ENAP Refinería Aconcagua)

Podemos visualizar estos datos desde el punto de vista Regional para evaluar localidad donde se vende el combustible y desde el punto de vista mensual para evaluar si hay estacionalidad en las ventas. La distribución de estos datos en función de las regiones se muestra en la Tabla 4.1.

Al observar la Tabla 4.1 y la Figura 4.4 se puede evidenciar que el consumo es estable en el tiempo en cada región. Los mayores consumos están en la Región Metropolitana y Los Lagos, siendo la primera la de mayor población del país y la segunda un territorio vasto con múltiples accidentes geográficos como Chiloé y sus archipiélagos, que necesitan precisamente de avionetas para su funcionamiento. Los datos del 2025, del primer trimestre, también muestran la misma tendencia. Esto último se puede observar en la tabla 4.2, donde se reafirma que la demanda es estable y tampoco tiene una estacionalidad definida.

4.2.4. El método estándar D909-22

El método ASTM D909-22 establece un procedimiento estandarizado para la determinación de las características antidetonantes de las gasolinas de aviación mediante el denominado Método de Supercarga. Su principal objetivo es evaluar la resistencia del combustible a la detonación bajo condiciones que simulan la operación real de motores aeronáuticos sobrealimentados, los cuales demandan elevados niveles de desempeño y seguridad.

Región	2022	2023	2024	2025⁸
Arica y Parinacota	9,3	9,1	8,3	2,9
Atacama	23,0	29,2	28,0	4,4
Tarapacá	228,7	108,4	63,8	12,1
Antofagasta	63,5	44,7	18,7	2,0
Coquimbo	78,3	157,4	112,6	27,0
Valparaíso	410,8	333,1	279,1	72,2
Metropolitana de Santiago	1504,7	1569,1	2033,8	548,0
Libertador General Bernardo O'Higgins	174,0	153,7	128,6	45,1
Maule	187,8	136,1	140,8	67,1
Ñuble	126,4	112,1	105,6	49,0
Biobío	205,2	223,0	280,5	87,8
La Araucanía	272,2	319,6	338,6	196,6
Los Ríos	83,0	72,0	63,0	37,0
Los Lagos	1183,0	1215,8	1196,9	335,0
Aysén del General Carlos Ibáñez del Campo	197,5	204,7	146,0	37,2
Total	4747,2	4687,9	4944,2	1523,4

Tabla 4.1: Valores por región y año

(Fuente: Superintendencia de Electricidad y Combustibles)

Total ventas por Mes	2022	2023	2024	2025
ENERO	493,6	506,5	607,2	553,7
FEBRERO	455,1	470,5	528,2	528,3
MARZO	526,0	395,5	411,2	441,3
ABRIL	372,8	399,4	421,3	
MAYO	356,2	288,2	341,1	
JUNIO	317,6	276,4	245,6	
JULIO	286,0	328,8	471,1	
AGOSTO	338,7	319,1	344,3	
SEPTIEMBRE	314,7	362,3	325,7	
OCTUBRE	370,6	377,2	402,1	
NOVIEMBRE	420,1	460,4	473,1	
DICIEMBRE	495,7	503,7	373,3	
Total general	4747,2	4687,9	4944,2	1523,4

Tabla 4.2: Total de ventas Gasolina de Aviación por mes y año

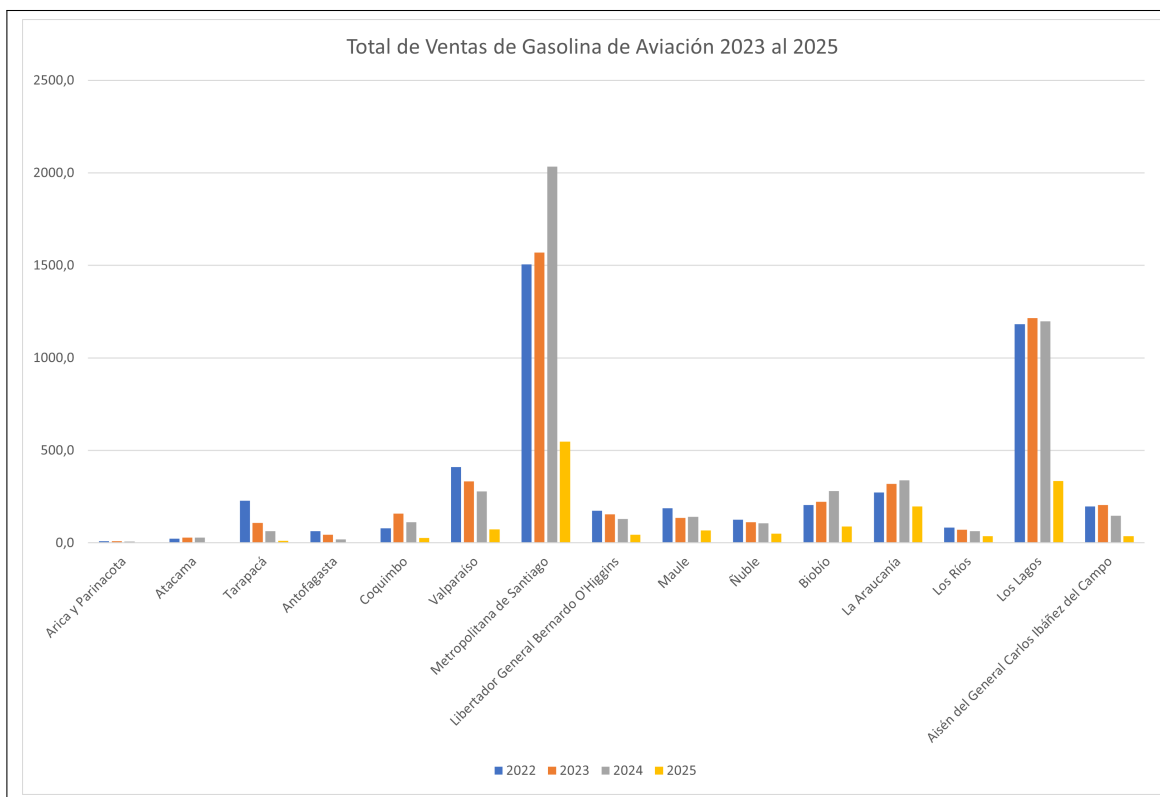


Figura 4.4: Total de Ventas de Gasolina de Aviación 2023 al 2025

(Fuente: Elaboración propia a partir de datos de la SEC)

El principio del ensayo se basa en la utilización de un motor monocilíndrico CFR (Cooperative Fuel Research), diseñado específicamente para pruebas de combustibles. Dicho motor opera bajo condiciones controladas de sobrealimentación, en las que parámetros como la presión de admisión, la relación airecombustible y la relación de compresión son cuidadosamente ajustados. La tendencia del combustible de ensayo a presentar fenómenos de detonación (knock) se compara con mezclas patrón de isoootano y tetráetilo de plomo (TEL), u otras referencias establecidas, permitiendo asignar un valor de Performance Number (PN).

El Performance Number constituye una extensión de la escala de octanos, en la cual el isoootano puro se define como PN 100. Dado que muchas gasolinas de aviación superan este valor, la escala se amplía utilizando aditivos antidetonantes que permiten alcanzar cifras en el rango de 110 a 130 PN. De este modo, el método ofrece una medida confiable de la capacidad antidetonante de los combustibles empleados en aviación, información crítica para garantizar la seguridad operativa de los motores y prevenir fallas asociadas a la detonación, como pérdida de potencia, sobrecalentamiento o daños mecánicos severos.

En síntesis, el ASTM D909-22 constituye una referencia esencial en la caracterización de gasolinas de aviación, ya que entrega un procedimiento reproducible y ampliamente aceptado para evaluar la calidad antidetonante en condiciones que representan fielmente las exigencias de los motores aeronáuticos modernos. En la normativa Chilena este dato se denomina *Detonancia mezcla rica*. El procedimiento en el laboratorio se rige por los siguiente pasos y se puede observar en la Figura 4.5 como se dibujan las curvas necesarias para determinar el valor de la propiedad:

1. Preparar el motor

- Se calienta con gasolina hasta que quede estable.

2. Probar la muestra

- Se coloca la gasolina que queremos medir.
- Se varía la mezcla de aire y combustible (más rica o más pobre).
- En cada ajuste, se mide cuánta potencia genera el motor y en qué punto empieza a detonar.
- Se toman al menos **6 mediciones** para trazar una curva.

3. Probar los combustibles de referencia

- Se repite el mismo proceso con dos gasolinas "patrón":
 - a) Una de menor resistencia (bajo octanaje).
 - b) Una de mayor resistencia (alto octanaje).

4. Comparar curvas

- Se dibujan las tres curvas (muestra + 2 referencias).
- Se ve dónde queda ubicada la muestra respecto a las referencias.

5. Calcular el resultado

- Si queda **bajo 100 octanos**, se informa como "número de octano".
- Si queda **sobre 100 octanos**, se expresa en equivalentes de TEL (tetraetilo de plomo) y se convierte a "Performance Number".

Hay otros métodos de análisis que son similares al ASTM D909-22, por ejemplo el método ASTM D2700 "Standard Test Method for Knock Characteristics of Motor Fuels by the Motor Method" se emplea principalmente para la determinación del número de octano motor (Motor Octane Number, MON) en gasolinas automotrices. Este procedimiento utiliza igualmente un motor CFR, pero opera bajo condiciones menos exigentes que las establecidas en el ensayo sobrealimentado. En el D2700, los parámetros de operación, tales como la presión de admisión, la temperatura de la mezcla aire-combustible y la relación de compresión, están estandarizados para representar condiciones de carga y velocidad más propias de motores terrestres, y no de motores aeronáuticos de alta potencia.

4.2.5. Enfoques predictivos en la industria

En los últimos años, el uso de modelos de machine learning en procesos industriales ha demostrado ser una herramienta poderosa para mejorar tiempos de respuesta, reducir costos y aumentar confiabilidad. En el caso de combustibles, estos modelos permiten aprovechar bases de datos históricas de laboratorio para construir predictores robustos que reduzcan la dependencia de ensayos costosos o poco disponibles, que pueden ser extrapolables a cualquier industria.

Tal como lo señala Nieto (2022) en su Tesis, hoy en día en la industria petrolera, tanto de upstream como downstream hay muchísimos casos donde el ML aplicado suele ser una alternativa válida para predecir comportamientos de yacimientos petroleros o predecir el comportamiento de ciertas mezclas de crudos a la entrada de una refinería.

Más allá del sector petrolero, existen múltiples aplicaciones transversales. En la industria química y farmacéutica, el ML se emplea para predecir rendimientos de reacciones, optimizar formulaciones y diseñar procesos más sostenibles. En manufactura avanzada, los modelos predictivos permiten anticipar fallas en equipos críticos (mantenimiento predictivo) y ajustar parámetros de producción en tiempo real. En el ámbito de la energía, se han utilizado para predecir el comportamiento de turbinas, redes eléctricas y procesos de distribución.

Los enfoques predictivos en la industria constituyen una línea de desarrollo con alto potencial de impacto, ya que combinan la disponibilidad creciente de datos con algoritmos cada vez más eficientes, entregando soluciones prácticas, escalables y alineadas con la transformación digital en curso en los sectores productivos.

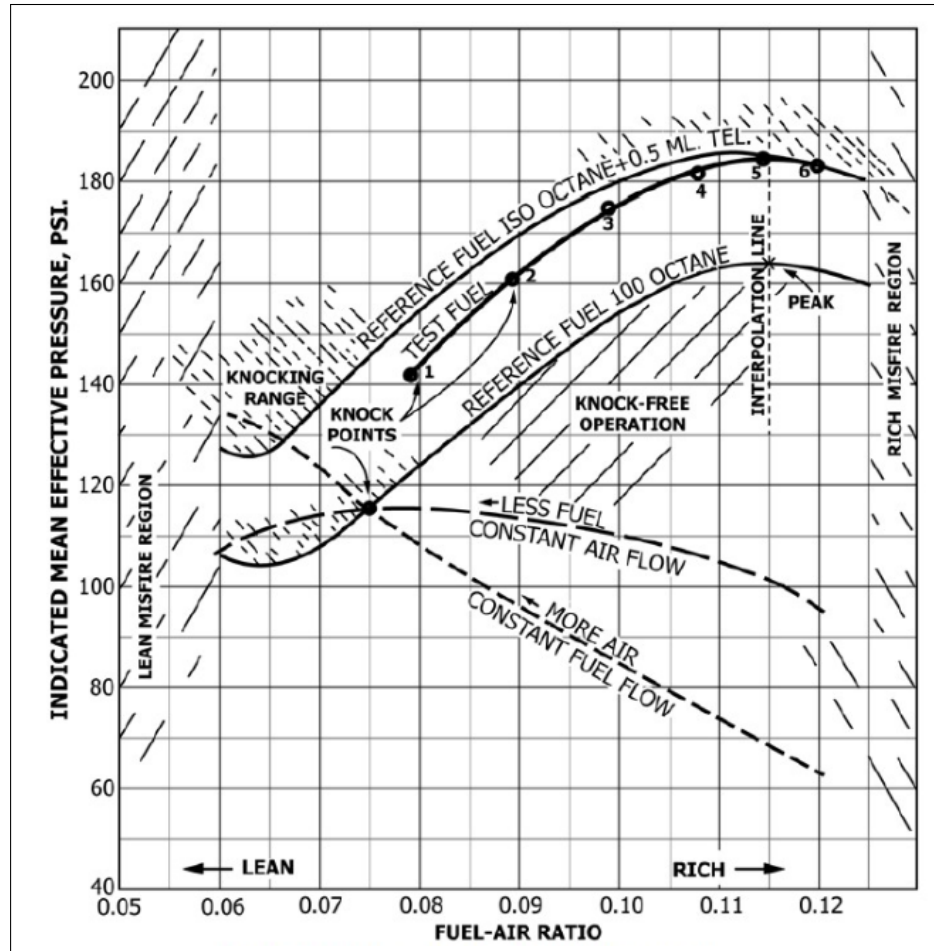


Figura 4.5: Desarrollo de curvas de potencia limitada por detonación

(Fuente: Norma ASTM D909-22)

4.3. Metodología utilizada

Este estudio adopta un enfoque cuantitativo y descriptivo. Se utilizarán técnicas de modelado predictivo y análisis estadístico aplicadas sobre un conjunto de datos estructurados para desarrollar el ML como servicio y luego para el desarrollo del modelo de negocios se utilizará un enfoque cuantitativo para determinar los costos y flujo de caja proyectado. La metodología contempla las siguientes fases:

1. **Análisis basado en Machine Learning:** Se crea el modelo piloto en base a la Detonancia mezcla rica en Gasolina de Aviación que tiene la siguiente secuencia de desarrollo:
 - Recolección de datos: Extracción de datos históricos desde el sistema LIMS. Se seleccionarán las muestras que contengan simultáneamente las propiedades requeridas para el análisis y el valor de detonancia medido por CFR.
 - Preparación de los datos: Revisión de consistencia, detección de valores atípicos, análisis de correlación entre variables y visualización gráfica de patrones.
 - Selección de variables: Aplicación de métodos estadísticos y de reducción de dimensionalidad (correlación) para determinar las variables más relevantes para la predicción.
 - Desarrollo de modelos: Entrenamiento de modelos de machine learning supervisado. Se considerarán al menos cuatro algoritmos distintos: regresión lineal múltiple, árbol de decisión, elastic net

y XGBoost. El entrenamiento se realizará utilizando validación cruzada y división entre conjunto de entrenamiento (70 %) y prueba (30 %) y luego entrenamiento (80 %) y prueba (20 %).

- Evaluación de desempeño: Se utilizarán métricas como el error cuadrático medio (MSE) y el error promedio para comparar los modelos y seleccionar el más adecuado.
 - Propuesta de implementación: Se propondrá un esquema conceptual de integración del modelo al flujo de trabajo del laboratorio, especificando condiciones mínimas de uso, alertas y validaciones cruzadas.
2. Modelo de Negocio: en base a experiencia en el desarrollo del piloto se paquetiza la metodología y el análisis como servicio. A partir de este se evalúan algunos supuestos para exportar el servicio a varios rubros y desarrollar otros modelos de ML.
- Análisis basado en ML como servicio: explicación de que elementos, y sus variantes, pueden ser utilizados para la prestación de estos servicios.
 - Breve mirada al mercado: revisión general de la oferta y demanda en Chile
 - Canvas: Desarrollo del modelo de negocios
 - Flujo de Caja estimado: Análisis de las variables para implementar el negocio.

4.4. Herramientas y entornos

Dentro de las herramientas que se utilizarán en esta tesina se encuentran principalmente herramientas de software herramientas para el desarrollo del modelo de negocio.

En términos de software tenemos las siguientes:

- **Lenguaje de programación:** Python 3.x
- **Librerías para ML:** pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost
- **Entorno:** Jupyter Notebook, Visual Studio Code
- **Visualización y análisis:** gráficos de dispersión, matrices de correlación, tablas de predicción vs valores reales.
- **Datos de Laboratorio:** Base de datos de LIMS SampleManager® y Winbliss®.

Para el desarrollo del modelo de negocio se utilizaron:

- **Business Model Canvas** para diseño estructural.
- Excel para flujo de caja y análisis financiero (VAN, TIR, punto de equilibrio).

4.5. Limitaciones

El modelo se construirá exclusivamente a partir de datos históricos. No se contemplan nuevas pruebas experimentales ni interferencias en procesos de producción. La validación regulatoria del modelo predictivo como sustituto del método D909-22 queda fuera del alcance de este desarrollo. La primera versión del modelo de negocio está limitada al análisis de laboratorio y su escalamiento a tiempo real en planta se plantea como línea futura. Los supuestos financieros se realizaron en escenarios estimados, sin comprometer cifras reales de mercado.

5 | Desarrollo Metodológico

5.1. Análisis basado en ML para la Detonancia Mezcla Rica en Gasolina de Aviación

5.1.1. Recolección de datos

La producción de Gasolina de Aviación es una campaña mensual en la Refinería. Esto significa que no es un proceso continuo y requiere de planificación y ajustes previos en la planta para que se puedan mezclar los componentes correctos y producirla. Esta campaña es efectiva entre 3 a 5 días, lo justo para llenar el tanque donde se almacena este producto. En términos de proceso, se requieren ajustes y circuitos de piping alineados para alimentar el tanque y en términos de control de calidad se ejecutan varios puntos de control: se toman muestras en el tanque y se analizan a medida que se va llenando. Dos tipos de muestras ocurren en este periodo

- Análisis Críticos, conjunto reducido de métodos ASTM para caracterizar el producto y estimar si cumplirá o no la especificación completa
- Certificación, conjunto completo de métodos ASTM que permitirá vender un producto que cumple con los estándares nacionales

Es en la Certificación donde se obtiene nuestra variable llamada Detonancia mezcla rica, por lo que en general, y observando la historia de la información disponible, hay un conjunto reducido de datos que se pueden analizar.

En este caso, la fuente de los datos fueron dos sistemas: Winbliss y SampleManager (Figura 5.1), este último en reemplazo del primero a partir de finales del 2019, por lo que la información a coleccionar tiene formatos diferentes, pero representan lo mismo. Se consideran al menos unos 20 años de datos.

Para obtener los datos que utilizaremos en nuestro modelo se seleccionarán las muestras que contengan simultáneamente las propiedades requeridas para el análisis y el valor de detonancia medido por el método ASTM D909-22. Otro método de obtención de los datos puede ser directamente desde la SEC, esto es debido a que lo vendido se reporta a este organismo regulador.

5.1.2. Preparación de los datos

Una vez que se obtiene la información, se realiza una limpieza inicial, siempre asegurando la calidad de los datos. Se revisan de forma general las columnas que corresponden a ambos sistemas y se hacen columnas equivalentes ya que en cada sistema tienen nombres distintos. Luego se filtra por muestras que contienen información completa en su fila y si aún persisten valores nulos, se seleccionan de manera general las muestras de certificación, ya que los otros tipos de muestras no contienen la variable a predecir⁹.

⁹Nota del autor: Esto puede conducir a un sesgo, ya que las muestras de certificación son las más cuidadas, pero si uno evalúa las condiciones de planta se dará cuenta que también son condiciones especiales para producir este tipo de gasolina, por lo tanto el entorno

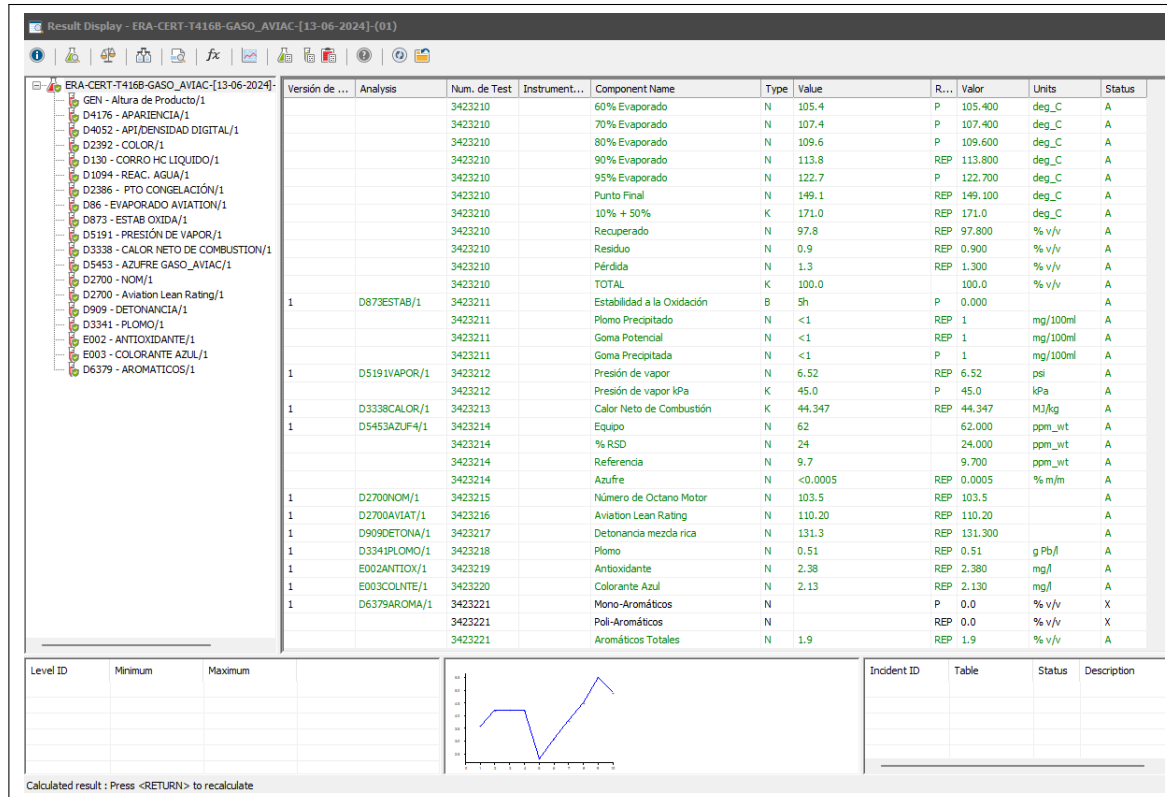


Figura 5.1: Resultados de muestra en SampleManager, LIMS

(Fuente: ENAP Refinería Aconcagua)

Se revisa la consistencia, se detectan valores atípicos, se analizan variables no numéricas y se realiza una correlación entre variables. Esto se puede revisar a través de un mapa de calor como en la siguiente Figura 5.2. Este mapa se puede ir modelando de forma iterativa hasta llegar a la selección de variables ideal. Como observación, en la columna final se puede observar como el Plomo es el valor que más pesa en la correlación con un 0,4, seguidos de la propiedad de Recuperado, Pérdida y Colorante Azul.

5.1.3. Selección de variables

Posteriormente se ejecutó un análisis exploratorio de datos (EDA), identificando correlaciones entre variables y descartando aquellas con alta redundancia o bajo impacto. En el listado inicial teníamos 53 columnas de características y quedan finalmente 23, que son las siguientes:

1. Numero de Secuencia
2. Punto de Muestreo
3. Densidad a 15 °C kg/m3 (D4052)
4. API Gravedad, Digital 60°F (D4052)
5. Punto inicial °C (D86)
6. 10 % °C (D86)
7. 20 % °C (D86)
8. 30 % °C (D86)

completo es resguardado.

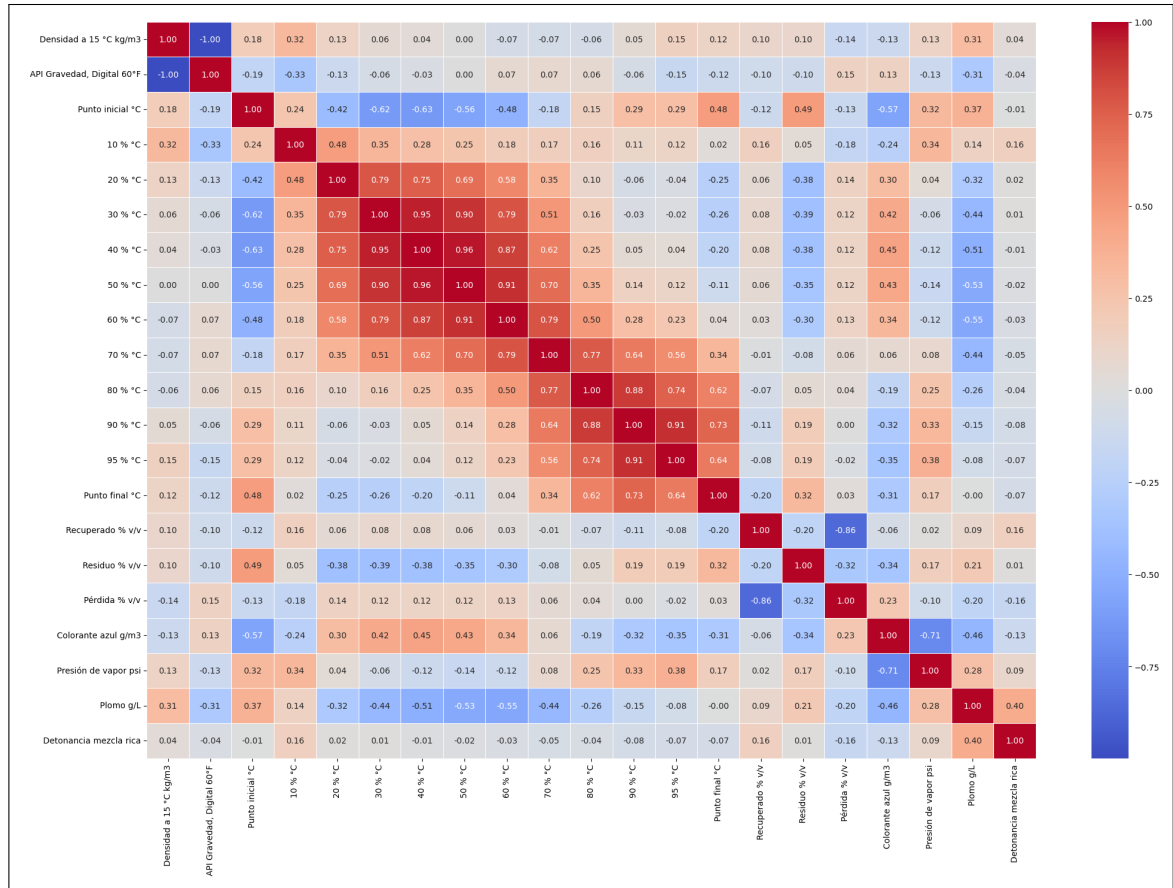


Figura 5.2: Mapa de Calor Resultados de variables analizadas

(Fuente: Elaboración propia en python)

9. 40 % °C (D86)
10. 50 % °C (D86)
11. 60 % °C (D86)
12. 70 % °C (D86)
13. 80 % °C (D86)
14. 90 % °C (D86)
15. 95 % °C (D86)
16. Punto final °C (D86)
17. Recuperado % v/v (D86)
18. Residuo % v/v (D86)
19. Pérdida % v/v (D86)
20. Colorante azul g/m³
21. Presión de vapor psi (D5191)
22. Plomo g/L (D3341)
23. Detonancia mezcla rica (D909-22)

Estas características son las más básicas para obtener la información. Además de ser elegidas por medio de variados filtros, la idea de tener un menor conjunto de datos para la estimación es mejor si se piensa en el resto de muestras descartadas por no tener el valor a predecir, que son las muestras críticas. Vale decir, en un futuro se podría agregar el análisis de detonancia a algunas muestras críticas y tener más datos de entrenamiento para mejorar el modelo, agregando valor al proceso de muestreo de tanque de gasolina de aviación.

5.1.4. Desarrollo de modelos y evaluación

En el desarrollo de modelos de machine learning es esencial seguir un proceso iterativo de selección y validación. Según Géron (2022), comparar distintos algoritmos bajo métricas comunes como MSE¹⁰ permite identificar el modelo más confiable y asegurar que sus resultados sean generalizables a nuevos datos.

Una vez que ya se seleccionaron los datos se evalúan cuatro modelos de machine learning:

- **Regresión Lineal:** Algoritmo básico que ajusta una recta o hiperplano para predecir valores, limitado para datos no lineales
- **Random Forest Regressor:** Conjunto de árboles de decisión que promedian sus resultados. Captura relaciones no lineales y es robusto, pero menos interpretable
- **Elastic Net:** Variante de regresión lineal con regularización L1 y L2. Reduce sobreajuste y selecciona variables, útil con datos correlacionados
- **XGBoost, Extreme Gradient boosting:** Algoritmo de boosting¹¹ basado en árboles, optimizado y muy preciso. Excelente rendimiento en problemas complejos, aunque requiere más ajuste de parámetros

El entrenamiento de los modelos de machine learning supervisado se realizará utilizando validación cruzada y división entre conjunto de entrenamiento (80 %) y prueba (20 %), además del 70 % y 30 %.

Se implementaron y compararon los cuatro modelos. El mejor modelo se seleccionó utilizando métricas de error estándar como el error cuadrático medio (MSE) y el error promedio para comparar los modelos y seleccionar el más adecuado.

5.1.5. Propuesta de implementación

Una vez que se selecciona el modelo se puede generar un archivo serializado tipo PKL que puede ser utilizado en la programación python de un sistema de integración. Este formato está diseñado específicamente para la biblioteca scikit-learn en Python.

En un esquema conceptual de integración del modelo predictivo al flujo de trabajo del laboratorio, se necesita que una vez que los datos de entrada hayan sido generados e ingresados, el sistema sea capaz de realizar una consulta al modelo y obtener el valor de predicción. Esto puede desarrollarse de múltiples formas, pero considerando que la complejidad pudiera incrementarse con el tiempo lo más conveniente es generar una API Rest que pueda devolver el dato de predicción respondiendo a una consulta de entrada, como en la Figura 5.3.

Por su simplicidad, se plantea desarrollar este modelo con el framework de FastAPI, disponible en Python. Este modelo es simple y escalable, por lo que puede ser utilizado para varias integraciones. Además, dependiendo de los sistemas donde se implemente, es posible llevarlo a la Cloud o implementarlo On-Premise, dependiendo de las políticas de cada organización.

¹⁰Mean Square Error, Error cuadrático medio

¹¹El boosting es una técnica donde se entrenan varios modelos débiles (como árboles pequeños) en secuencia. Cada nuevo modelo se enfoca en los errores que cometió el anterior, y al final se combinan todos para formar un modelo mucho más fuerte y preciso.

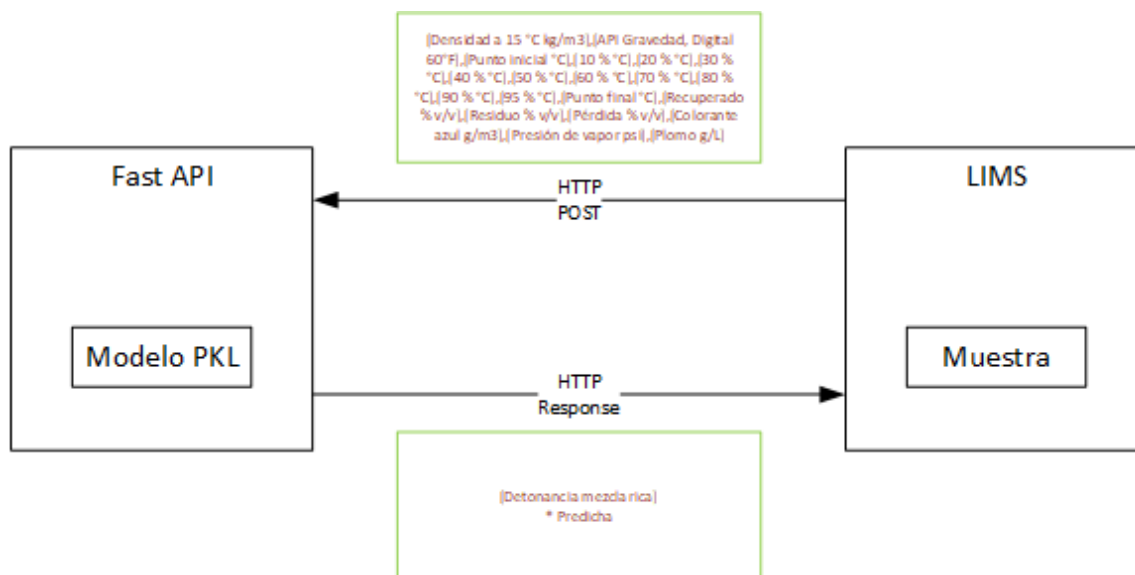


Figura 5.3: Modelo de Implementación Técnica

(Fuente: Elaboración propia)

5.2. Modelo de Negocio

5.2.1. Análisis basado en ML como servicio

El desarrollo del modelo predictivo para la Detonancia Mezcla Rica constituye un caso piloto que valida la factibilidad técnica de aplicar machine learning a análisis de laboratorio. No obstante, el valor principal de este trabajo se encuentra en su potencial de transformar esto en un servicio replicable, es decir, ofrecer Análisis basados en ML como Servicio (MLaaS) para la industria de combustibles y eventualmente otros sectores.

Este concepto se fundamenta en tres ejes principales:

- **Paquetización del proceso:** El desarrollo del modelo se estructura en fases estandarizadas como recolección de datos, limpieza, entrenamiento, validación y despliegue. Esta metodología puede aplicarse de forma sistemática a distintos análisis de laboratorio como por ejemplo, la predicción de presión de vapor, azufre, densidad, octanaje. Cada análisis se convierte en un módulo de servicio independiente, que puede ofrecerse de manera incremental a los clientes.
- **Entrega como servicio (MLaaS):** Los modelos se implementan en una plataforma digital conectada a los sistemas de laboratorio (LIMS) o a entornos de planta (Historizador, DCS). El cliente accede a los resultados mediante interfaces web o APIs, sin necesidad de gestionar infraestructura propia. Se establece un esquema de suscripción mensual o licenciamiento anual que incluye uso del software, actualizaciones y soporte técnico.
- **Escalabilidad hacia planta:** Mientras que el piloto se limita al entorno batch¹² de laboratorio, el mismo enfoque puede evolucionar hacia la predicción en tiempo real, integrando datos de sensores y sistemas de control industrial. Esto abre un mercado de mayor escala, donde los servicios de ML no solo sustituyen análisis costosos, sino que también optimizan la operación continua de refinerías y plantas energéticas.

En este sentido, el negocio no radica en comercializar un único modelo, sino en ofrecer un portafolio de modelos predictivos o *Analítica Predictiva* como servicio, bajo un esquema de suscripción, con posibilidad

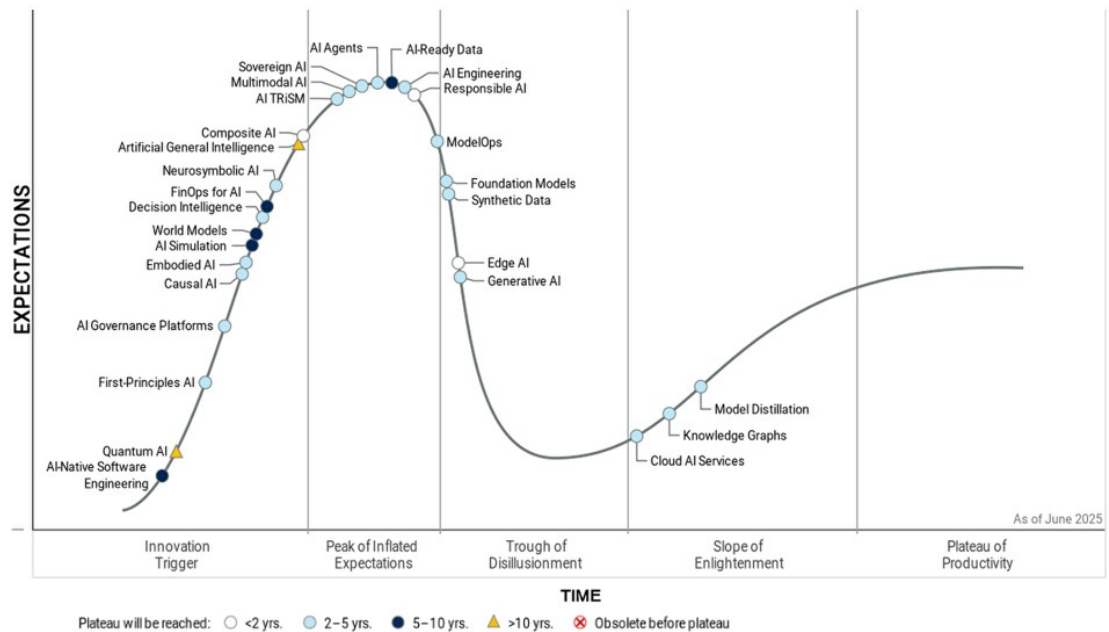
¹²se refiere a que se toman muestras manuales y se analizan posteriormente, no es en "línea." proceso

de escalamiento a distintas industrias y países.

La propuesta más amplia radica en la Analítica Predictiva como Servicio para la industria energética, con potencial de convertirse en un negocio sostenible y escalable en el tiempo.

5.2.2. Breve mirada al Mercado

El mercado para servicios de análisis predictivo en la industria de combustibles se encuentra en una etapa mixta de desarrollo en Chile y Latinoamérica, lo que representa tanto un desafío como una oportunidad estratégica. En particular la industria de Oil and Gas (hoy en día *Energía*) a pesar de que hay esfuerzos para aplicar ML en planificación y monitoreo de activos, sigue estando un poco más atrás respecto de otras industrias como retail o minería. En términos del *Hype Cycle* que genera [Gartner, Inc. \(2025\)](#) (Figura 5.4) nos encontramos en distintos periodos entre implementaciones de ciertas tecnologías, con poca visibilidad del valor que genera, y el desarrollo de políticas y gobernanza de Inteligencia Artificial.



Gartner

Figura 5.4: Hype Cycle Inteligencia Artificial 2025

(Fuente: Gartner)

La forma en que operan los laboratorios reviste una necesidad evidente de contar con alternativas que reduzcan la dependencia de equipos obsoletos, agilicen los tiempos de análisis y mejoren la eficiencia operativa de los mismos en pos de entregar mejores resultados a planta.

En el contexto latinoamericano, países como Brasil, Colombia, Ecuador, Perú y Argentina presentan un mercado con características similares:

- Producción concentrada en unas pocas refinerías, que no compiten entre si.
- Regulación estricta sobre calidad de combustibles.
- Dependencia de laboratorios centralizados con equipos analíticos costosos y limitados en disponibilidad.
- Avance incipiente en la adopción de herramientas digitales predictivas para control de calidad.

Según Zoting (2025), para el 2034 el mercado de Machine Learning valdrá unos 1.407,65 mil millones de dólares (Figura 5.5), teniendo en cuenta que latinoamérica representa el 8 % (Figura 5.6) de ese mercado el 2024, si se mantiene al 2034 serán unos 112,61 MMUSD.

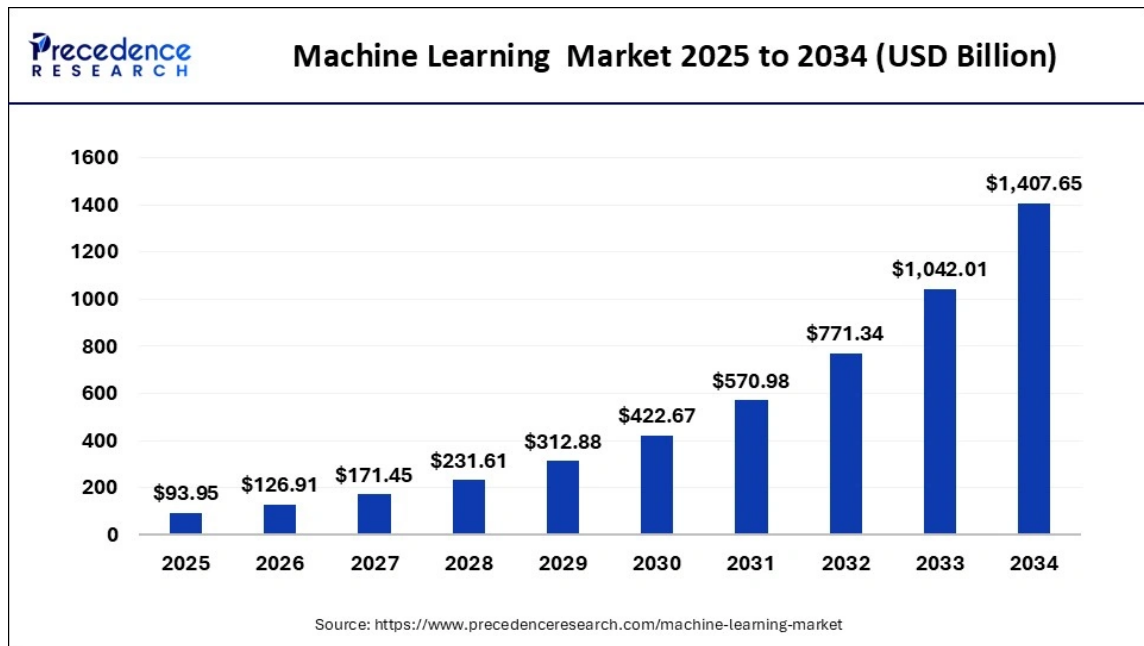


Figura 5.5: Mercado Global de Machine Learning 2025 al 2034

(Fuente: Precedence Research)

En síntesis, el mercado chileno constituye un prototipo natural para validar la propuesta, mientras que el mercado latinoamericano representa un espacio de expansión progresiva, donde la convergencia de innovación tecnológica, regulación energética y necesidad de eficiencia abre la puerta a un negocio con potencial regional.

5.2.3. Canvas

Según Osterwalder y Pigneur (2010), un Canvas es una herramienta visual llamada Business Model Canvas (o lienzo del modelo de negocio), que describe, diseña y analiza modelos de negocio utilizando una plantilla gráfica con nueve bloques interconectados. Esta herramienta ayuda a las empresas a definir de forma clara y sencilla sus propuestas de valor, su infraestructura, su mercado y su viabilidad económica, facilitando la innovación y la comunicación del negocio. A continuación se presenta un canvas enfocado en diseñar un modelo de negocio que permita la materialización de una empresa que ofrece servicios de Machine Learning para la industria de laboratorios, extensible a otros nichos (Figura 5.7).

5.2.3.1. Segmentos de Cliente

Los clientes para este servicio son todos aquellos laboratorios de combustibles, refinerías, industrias de extracción, organismos reguladores, industrias de la región.

5.2.3.2. Propuesta de Valor

La propuesta de valor de este negocio es acercar el Machine Learning a la industria mediante una plataforma paquetizada y escalable, capaz de reducir costos y tiempos en análisis críticos de laboratorio,

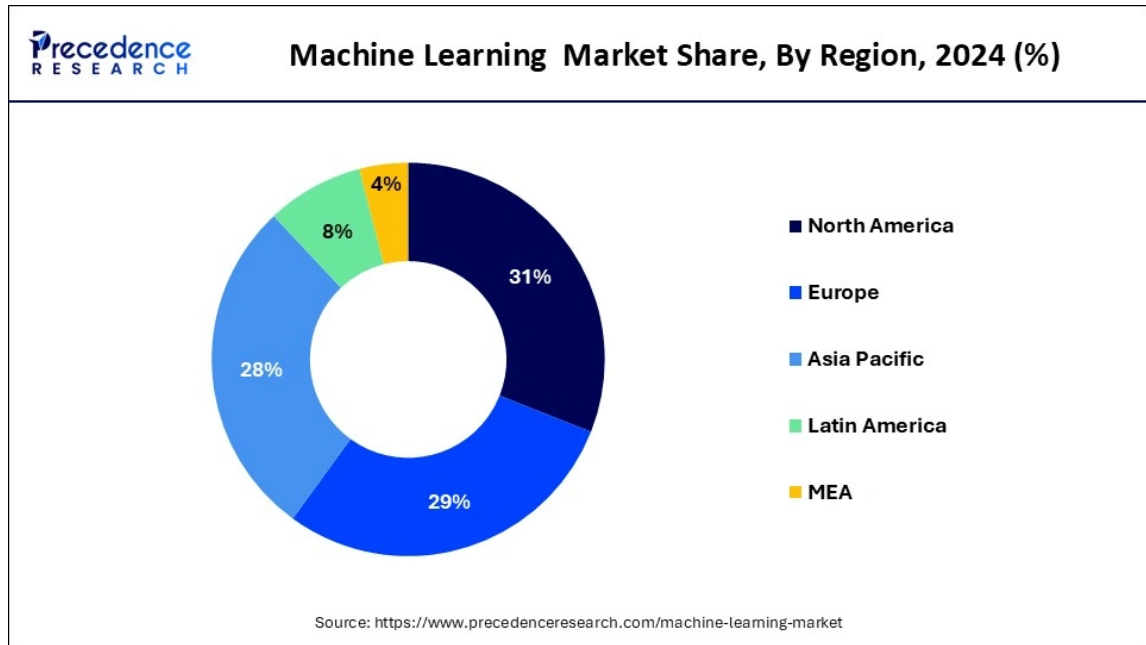


Figura 5.6: Mercado por región de Machine Learning 2024

(Fuente: Precedence Research)



Figura 5.7: Canvas Bussines Model

(Fuente: Elaboración propia)

sustituir parcialmente métodos tradicionales obsoletos y evolucionar hacia soluciones de monitoreo en línea en plantas de producción.

5.2.3.3. Canales de Distribución y Comunicación

Al ser un servicio tecnológico, su distribución radica en SaaS/MLaaS en la nube u On-Premise como fuente de cálculo para las predicciones y su integración con otros sistemas del Cliente.

5.2.3.4. Relación con el Cliente

Se contemplan los siguientes tipos de relacionamiento con el cliente:

- Contratos de suscripción
- Contratos de Implementación de proyectos
- Soporte técnico
- Capacitación

5.2.3.5. Fuentes de Ingreso

Para el desarrollo de este modelo de negocio se contemplan las siguientes fuentes de ingreso

- Suscripción anual o mensual por cada modelo
- Servicios de implementación de nuevos modelos
- Capacitación
- Consultoría

Para el primer año se espera realizar unos 6 proyectos que pasarían a continuidad operacional en formato suscripción y al segundo año se proyectan 19 proyectos adicionales. La idea es finalizar el año 5 con 85 suscripciones, que podría ser considerado lo aceptable para subsistir sin proyectos en el entendido de que el precio por suscripción está dentro de mercado.

5.2.3.6. Recursos Clave

Se necesitan los siguientes recursos claves para implementar tanto el modelo de negocio como cualquier proyecto de ML:

- Infraestructura Multi-Cloud
- Personal: Data Science y Programación
- Datos Históricos de planta y laboratorio

Se externalizará todo lo que sea Contabilidad y RRHH, modelo de contratación de servicios. Se considera arriendo de un cowork para los 2 primeros años y en el futuro se podría analizar un arriendo más amplio.

5.2.3.7. Actividades Clave

Dentro de las actividades claves tenemos dos tipos: Actividades técnicas y comerciales:

- Técnicas: entrenamiento de modelos, validación, soporte y actualizaciones, elaboración de proyectos.
- Comerciales: Actividades de ventas, workshops, email marketing, asistencia a ferias tecnológicas para networking.

5.2.3.8. Socios Clave

Nuestros socios estratégicos son universidades, proveedores cloud, jefes de laboratorios, directores de TI.

5.2.3.9. Estructura de Costos

Se planea tener una estructura de costos mínima, trabajo presencial y remoto al menos los 2 a 3 años iniciales con un equipo conformado por:

- Jefe de Proyectos
- Científico de Datos
- Programador

Además requerimos infraestructura básica para la operación de la empresa, cowork, telemática, notebooks y servicios cloud.

5.2.4. Flujo de caja

El flujo de caja proyectado del negocio se construye a partir de un modelo híbrido que combina ingresos recurrentes, en formato suscripción, y extraordinarios, en formato de proyectos y consultoría, con el objetivo de alcanzar un equilibrio financiero sostenible en el mediano plazo.

En primer lugar, la principal fuente de ingresos proviene del modelo de suscripción, el cual asegura flujos periódicos y predecibles. Esta modalidad permite proyectar con mayor certeza el crecimiento de la base de clientes en un horizonte de cinco años, estableciendo así la estabilidad necesaria para cubrir costos operativos y reinvertir en el desarrollo del servicio. En la Tabla 5.1 se muestra el cálculo de las suscripciones bajo un escenario realista.

En paralelo, se incorporan ingresos complementarios mediante la venta de proyectos personalizados, los cuales son valorizados en función de horas de trabajo de distintos roles profesionales (Sueldos Tabla 5.1). Esta línea de negocio aporta flexibilidad y capacidad de diversificación hacia otros sectores industriales, disminuyendo el riesgo de depender únicamente de las suscripciones, ya que este ítem se construye en el largo plazo.

En cuanto a los egresos (Tabla 5.2), se consideró una estructura de costos ajustada, donde el mayor componente corresponde a remuneraciones y herramientas tecnológicas. La estrategia contempla mantener una estructura liviana en los primeros años, incorporando un gerente a partir de finales del primer año para reforzar la gestión y venta, y postergando la inversión en infraestructura o inmuebles con el fin de no comprometer liquidez innecesaria.

	Año 1	Año 2	Año 3	Año 4	Año 5
Valor UF	39.485,65	42.249,65	44.362,13	46.580,23	48.909,25
Suscripciones (MLaaS)	6	19	35	60	85
Precio x Suscripción	159.917	181.634	208.879	240.211	276.243
Cant. Proyectos	6	19	10	25	25
Cant. HH J. Proyectos	210	630	300	750	750
Cant. HH Analista Datos	1.080	3.780	1.800	4.500	4.500
Cant. HH Programador	600	2.100	1.000	2.500	2.500
Ventas por Clientes Web	3,62	35,42	87,73	172,95	281,77
Ventas por Proyectos	124,14	385,62	158,11	456,53	527,30
Sueldos	78,65	137,12	153,48	243,73	350,51
Compras con IVA	1,06	1,06	1,17	1,29	1,42
INGRESOS					
Ventas c/n IVA	152,03	501,03	292,54	749,09	962,78
Cobro (0)	71,70	347,85	204,78	524,36	673,95
Cobro (30)	23,26	147,81	87,76	224,73	288,84
Total Ingresos	94,96	495,66	292,54	749,09	962,78

Tabla 5.1: Total de Ingresos (Millones de pesos)

Bajo este esquema, el flujo de caja muestra un inicio con resultados negativos, propios de la inversión inicial y del lento despegue comercial, pero que progresivamente se corrige a medida que crece la base de suscriptores y se concretan proyectos específicos. El punto de equilibrio se proyecta hacia la mitad del periodo analizado, alcanzando en los últimos años una posición financiera positiva que permite pensar en la sostenibilidad y eventual expansión del negocio. Se considera que hay una caja inicial de 40M CLP.

	Año 1	Año 2	Año 3	Año 4	Año 5
EGRESOS					
Proveedor (0)	0,80	5,99	6,58	7,24	7,97
Proveedor (30)	0,58	0,69	0,75	0,83	0,91
Salarios	78,65	139,52	153,48	243,73	350,51
IVA	15,53	78,80	79,46	0,00	118,58
PPM	1,72	8,35	8,42	4,92	12,59
Marketing	2,40	2,40	2,64	2,90	3,19
Gastos Generales	8,40	9,60	10,56	11,62	12,78
Adquisición Materiales	4,00	2,00	4,40	4,84	5,32
Total Egresos	112,07	247,35	266,29	276,08	511,86
Caja Inicial	40,00	22,89	271,20	297,45	770,46
Sup/Def	-17,11	248,31	26,25	473,00	450,93
Caja Final	22,89	271,20	297,45	770,46	1.221,38

Tabla 5.2: Total de Egresos y Diferencia de caja (Millones de pesos)

6 | Resultados Obtenidos

6.1. Modelo de ML

Los modelos que describimos en 5.1.4 son finalmente los modelos que se entrenan con un conjunto de datos y se obtienen métricas que nos permiten diferenciar la performance de cada modelo y decidir cual de ellos se ajusta mejor a lo requerido.

Con el siguiente código en Python (Código 6.1) se define una función que es capaz de obtener los valores de MSE, o Error cuadrático medio y el error promedio.

Código 6.1: Función de evaluación modelo

```

1 # Función para evaluar un modelo y calcular MSE y Error Promedio
2 def evaluar_modelo(modelo, X_train, X_test, y_train, y_test):
3     modelo.fit(X_train, y_train)
4     y_pred = modelo.predict(X_test)
5     mse = mean_squared_error(y_test, y_pred)
6     error_promedio = mean_absolute_percentage_error(y_test, y_pred) * 100
7     return mse, error_promedio

```

Luego con el siguiente código (6.2) se evalúa cada modelo y se obtienen las métricas para comparar, resultando el modelo *Random Forest Regressor* con la puntuación más baja de MSE y de error promedio, según la Tabla 6.1. En esta tabla se muestran 2 conjuntos de datos realizados con un 70 % y 80 % de datos de entrenamiento.

Modelo	Entrenamiento (70 %)		Entrenamiento (80 %)	
	MSE promedio	Error promedio	MSE promedio	Error promedio
Regresión Lineal	2.888	1.01 %	2.757	0.95 %
Random Forest Regressor	2.567	0.93 %	2.250	0.81 %
Elastic Net	2.765	0.99 %	2.900	0.99 %
XGBoost	2.851	0.97 %	3.025	0.98 %

Tabla 6.1: Comparación de desempeño de los modelos en entrenamiento (70 % de los datos) y test (30 % de los datos).

En la comparación de los distintos modelos evaluados, el Random Forest Regressor presentó el mejor desempeño global. Este modelo obtuvo el MSE promedio más bajo (2.567/2.250) y, al mismo tiempo, el menor error promedio (0.93 %/0.81 %), en ambos casos, lo que indica que sus predicciones fueron más precisas y consistentes respecto de los valores reales. La superioridad del Random Forest se explica por su capacidad de combinar múltiples árboles de decisión mediante la técnica de *bagging*, lo que reduce la varianza y mejora la robustez frente a datos ruidosos o no lineales. En contraste, la Regresión Lineal mostró limitaciones al no capturar adecuadamente relaciones complejas entre variables, mientras que Elastic Net y XGBoost, aunque potentes en otros contextos, no alcanzaron el mismo nivel de ajuste en este caso particular. Por lo tanto, se concluye que el Random Forest Regressor constituye la mejor alternativa para el problema

analizado, ya que ofrece un equilibrio óptimo entre bajo error cuadrático medio y baja tasa de error relativo, asegurando mayor capacidad de generalización.

Estos resultados iniciales, luego tienen que ser comparados con datos reales, vale decir, tomar el restante de datos de test para realizar las comparativas finales del modelo y obtener nuevamente los valores de MSE y Error promedio.

Código 6.2: Métricas por modelo

```

1 # 1. Regresión Lineal
2 modelo_lr = LinearRegression()
3 mse_lr, error_lr = evaluar_modelo(modelo_lr, X_train, X_test, y_train, y_test)
4
5 # 2. Random Forest Regressor
6 modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)
7 mse_rf, error_rf = evaluar_modelo(modelo_rf, X_train, X_test, y_train, y_test)
8
9 # 3. Elastic Net
10 modelo_en = ElasticNet(random_state=42)
11 mse_en, error_en = evaluar_modelo(modelo_en, X_train, X_test, y_train, y_test)
12
13 # 4. XGBoost
14 modelo_xgb = xgb.XGBRegressor(random_state=42)
15 mse_xgb, error_xgb = evaluar_modelo(modelo_xgb, X_train, X_test, y_train,
    y_test)

```

Como ya elegimos en este código el Random Forest, volvemos a entrenar y a evaluar, esta vez solamente con este modelo con el Código 6.3, en la línea 18 utilizamos los datos de Test que se seleccionaron en la línea 9 (mediante split aleatorio). En las líneas finales se muestra que el modelo se puede paquetizar en un archivo PKL que puede ser utilizado como modelo preentrenado para cualquier aplicación. En nuestro caso, tomamos este archivo y lo utilizamos para crear un API Rest con la librería FastAPI, entonces queda un modelo accesible a cualquier sistema.

Código 6.3: Evaluar Modelo Random Forest

```

1 from sklearn.model_selection import train_test_split
2 from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error
3
4 # Asegúrate de que X e y estén correctamente definidos (ya lo has hecho en los
    pasos anteriores)
5 # X = df_limpio.drop('Detonancia mezcla rica', axis=1)
6 # y = df_limpio['Detonancia mezcla rica']
7
8 # Dividir los datos en conjunto de entrenamiento y prueba
9 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
    random_state=42)
10
11 # Crear el modelo RandomForest
12 modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)
13
14 # Entrenar el modelo
15 modelo_rf.fit(X_train, y_train)
16
17 # Hacer predicciones en el conjunto de prueba
18 y_pred = modelo_rf.predict(X_test)
19
20 # Calcular el MSE y el error promedio
21 mse_rf = mean_squared_error(y_test, y_pred)
22 error_rf = mean_absolute_percentage_error(y_test, y_pred) * 100
23

```

```

24 # Mostrar los resultados
25 print(f"Resultado_RandomForrest\nMSE_promedio:_{mse_rf:.3f}\nError_promedio:_{
    error_rf:.2f}%\n")
26
27 y_pred
28 modelo_rf
29
30 import joblib
31
32 # Serializar el modelo Random Forrest en un archivo .pkl
33 archivo_modelo = "modelo_rf.pkl"
34 joblib.dump(modelo_rf, archivo_modelo)
35
36 print(f"El_modelo_RandomForrest_ha_sido_guardado_en:_{archivo_modelo}")

```

La evaluación final se toma un 80 % de los datos para el entrenamiento final y se vuelve a evaluar su capacidad. En ambos casos los resultados dan exactos, por el tamaño del set de datos en general.

Conjunto	MSE promedio	Error promedio
Entrenamiento	2.250	0.81 %
Test	2.250	0.81 %

Tabla 6.2: Comparación del desempeño del modelo Random Forest en entrenamiento y test

Si tomamos una muestra de esta información veríamos cuan cercano o alejada está la estimación. En la siguiente Tabla 6.3 se muestran 20 datos de test aleatorios con su respectivo valor y su predicción.

Índice	Detonancia mezcla rica	Predicción
66	132.0	133.336
45	134.0	134.910
9	132.0	133.200
73	134.0	133.502
136	132.0	134.987
104	130.0	132.738
205	134.0	133.496
247	130.9	132.225
125	136.0	134.532
211	132.5	132.475
84	132.0	133.481
33	134.0	134.070
254	133.1	132.866
68	136.0	134.021
46	136.0	134.375
108	131.0	131.710
6	133.0	133.069
30	133.0	132.945
101	134.0	133.251
55	134.0	132.132

Tabla 6.3: Comparación entre valores reales de detonancia mezcla rica y predicciones del modelo.

Lo logrado con este modelo se puede resumir en los siguientes puntos:

- El modelo Random Forest presentó el mejor desempeño con un Error de 0.81 % y un MSE de 2,250.
- Se logró reducir el tiempo de estimación de la detonancia de varias horas a segundos.

- Se identificaron 7 variables clave que explican más del 85 % de la variación de la variable objetivo.
- El modelo fue robusto ante datos nuevos dentro de los rangos esperados.
- Este modelo puede ser utilizado para las muestras críticas, ya que coinciden la cantidad de parámetros de *entrada*¹³.

6.2. Modelo de Negocio

Al revisar el balance general (Tabla 6.4) que se obtiene del flujo de caja proyectado, se pueden ver tres cosas importantes.

Primero, en los activos lo más relevante es la caja que se va generando año a año y algunos equipos que se compran para la operación. Esos equipos se consideran con depreciación acelerada, lo que significa que pierden valor más rápido en los papeles, pero refleja mejor la realidad porque en el corto plazo hay que pensar en renovaciones o actualizaciones.

Segundo, en los pasivos la situación es bastante baja, lo que es positivo porque el proyecto no queda muy expuesto a deudas. Esto ayuda a que la operación no dependa de créditos ni de grandes compromisos financieros que puedan presionar los resultados.

Tercero, está el capital de trabajo, que básicamente corresponde al aporte inicial que se hace para cubrir sueldos, servicios y los gastos más básicos de operación. Este capital permite darle oxígeno al proyecto mientras las suscripciones y proyectos empiezan a crecer y consolidarse.

En resumen, el balance general muestra un negocio que parte con una estructura liviana: activos moderados, pocos pasivos y un capital de trabajo bien enfocado. Eso le da solidez en los primeros años y, a medida que los ingresos van aumentando, abre la posibilidad de reinvertir y proyectar el crecimiento a futuro.

¹³Es necesario aclarar que hoy en día estos datos no existen, pero se podría establecer un caso en donde se realice el análisis de la detonancia en las muestras críticas para reentrenar el modelo con estos nuevos datos.

	Inicial	Año1	Año2	Año3	Año4	Año5
Balance General						
Caja		22,89	271,20	297,45	770,46	1.221,38
Inventario		0,00	0,00	0,00	0,00	0,00
Cuentas por Cobrar (Cx C)		0,00	62,44	62,44	62,44	62,44
PPM		1,72	10,07	18,49	23,41	36,00
Maquinarias		3,36	3,36	3,36	3,36	3,36
Depreciación Ac.		-0,67	-1,34	-2,69	-4,03	-5,38
Total Activos		27,30	345,73	379,06	855,64	1.317,81
Pasivos y Patrimonio						
Proveedores		-0,12	-5,52	-11,47	-18,01	-25,20
IVA		21,16	79,46	0,00	118,58	152,60
Línea de Crédito		0,00	0,00	0,00	0,00	0,00
ImptoxPagar		8,97	69,60	84,41	166,34	263,17
Capital Aportado		40,00	40,00	40,00	40,00	40,00
Utilidad del Ej.		30,89	239,73	290,75	572,96	906,48
Total Pasivos y Patrimonio		100,91	423,27	403,69	879,88	1.337,05
Estado de Resultado						
Ventas		127,76	548,80	794,63	1.424,12	2.233,18
Costo de Ventas		-0,18	-1,95	-16,58	-25,23	-39,31
Sueldos		-78,65	-218,17	-371,65	-615,38	-965,88
G. Generales		-8,40	-18,00	-28,56	-40,18	-52,95
Depreciación		-0,67	-1,34	-2,69	-4,03	-5,38
G. Financieros		0,00	0,00	0,00	0,00	0,00
Utilidad Bruta		39,86	309,33	375,16	739,30	1.169,65
Impto		-8,97	-69,60	-84,41	-166,34	-263,17
Utilidad Neta		30,89	239,73	290,75	572,96	906,48
EBITDA	-40,00	40,53	310,67	377,85	743,34	1.175,03
Inversión	-40,00					
Flujo de Caja Libre (FCL)	-40,00	30,89	239,73	290,75	572,96	906,48

Tabla 6.4: Balance General Realista (en Millones de pesos)

Como un ejercicio adicional para este modelo de negocio, determinaremos el Valor de la Idea en distintos escenarios basados en el Flujo de Caja Libre, donde propongo diversas situaciones y tasas. En particular las tasas utilizadas están en la Tabla 6.5, y en general se castiga fuertemente los primeros años. En estas estimaciones podemos observar que el *Valor de la IDEA* oscila ente 250 y 876 millones de pesos, Tabla 6.6.

	Año 1	Año 2	Año 3	Año 4	Año 5	Situación al Año 5
r Realista	50 %	45 %	30 %	25 %	25 %	Se han desarrollado 85 proyectos/suscripciones
r Pesimista	50 %	50 %	30 %	25 %	25 %	Se han desarrollado 50 proyectos/suscripciones
r Optimista	50 %	40 %	30 %	25 %	20 %	Se han desarrollado 110 proyectos/suscripciones

Tabla 6.5: Tasas y situaciones proyectadas bajo escenarios Real, Pesimista y Optimista

En el caso de la valorización de la empresa, podemos formular su valor en términos de su EBITDA

que suele utilizarse en este concepto. Para las tres situaciones, obtenemos un EBITDA de Valor Presente y se calcula el valor de la empresa, que es 5,4 veces el EBITDA de Valor presente, dando un valor que oscila ente \$2.046 y \$6.406 millones de pesos.

	Realista	Pesimista	Optimista
VAN*	\$601	\$290	\$916
VAN	\$561	\$250	\$876
Idea	\$561	\$250	\$876
Inversión	\$40	\$40	\$40

Tabla 6.6: Escenarios de evaluación: Real, Pesimista y Optimista (Millones de pesos)

	Real	Pesimista	Optimista
EBITDA Valor Presente	\$780	\$379	\$1.186
Valor Empresa/EBITDA	5,4	5,4	5,4
Valor Empresa	\$4.211	\$2.046	\$6.406

Tabla 6.7: Valoración de la empresa en distintos escenarios (Millones de pesos)

7 | Conclusiones

7.1. Declaraciones Generales

El desarrollo del modelo predictivo demostró que es factible estimar la Detonancia de Mezcla Rica en gasolinas de aviación con alta precisión, reduciendo tiempos y costos frente al método ASTM D909-22. El modelo de machine learning Random Forest Regressor se consolidó como el algoritmo más robusto, validando la hipótesis central de que los datos históricos de laboratorio constituyen una base sólida para aplicaciones de machine learning en la industria de combustibles.

Desde el punto de vista estratégico, la investigación evidenció que el verdadero valor no reside únicamente en este caso piloto, sino en su capacidad de ser replicado y paquetizado hacia otros análisis críticos de laboratorio y, posteriormente, hacia predicciones en línea en plantas de producción. Esto abre la posibilidad de construir un portafolio de servicios MLaaS para la industria energética y más allá.

No obstante, se identifican limitaciones relevantes: la validación regulatoria para el uso de estos modelos frente a la SEC aún queda fuera del alcance de este estudio; el tamaño reducido del dataset histórico podría influir en la generalización; y el modelo de negocio requiere mayor maduración en torno a la aceptación del mercado y la gestión del cambio en industrias altamente reguladas.

Como proyección futura, se recomiendan las siguientes consideraciones:

1. Ampliar la base de datos incorporando más muestras críticas y experimentales.
2. Iniciar procesos formales de validación del método de cálculo por ML con la SEC.
3. Desarrollar pilotos de integración en tiempo real en refinerías con otros análisis más relevantes al proceso.
4. Expandir el modelo de negocio hacia otros sectores industriales que comparten desafíos similares en análisis de laboratorio.
5. Elaborar una base de conocimiento analítico y modelos pre-entrenados que permita disponibilizar mejores modelos a otras compañías de la región.

En conclusión, esta tesina constituye un aporte concreto a la digitalización y modernización de la industria de combustibles, alineando innovación tecnológica y emprendimiento para responder a problemas críticos de calidad, eficiencia y sostenibilidad en mercados altamente exigentes.

7.2. Reflexión Final

La incorporación de modelos de machine learning en entornos industriales representa una oportunidad concreta para modernizar procesos, aumentar eficiencia y reducir riesgos operativos, aún cuando hay mucha competencia en el mercado. Lo ideal es ir especializando el conocimiento para crear una empresa que sea capaz de diferenciarse de otras.

Bibliografía

- ASTM International (2022). ASTM D909-22: Standard Test Method for Supercharge Rating of Spark-Ignition Aviation Gasoline. <https://www.astm.org/d0909-22.html>. Accessed: 2025-05-15. 1
- Dirección General de Aeronáutica Civil (2019). 1942, Se crea el Registro Nacional de Matrículas de Aeronaves. <https://www.dgac.gob.cl/1942-se-crea-el-registro-nacional-de-matriculas-de-aeronaves/>. Accessed: 2025-08-15. 4.2.1
- ENAP (2023). Procedimiento para la determinación del desempeño supercarga de gasolina de aviación de encendido por chispa. Documento interno, código ERA-LAB-PRA-D909-R2. Aprobado 14-02-2023. 1
- Gartner, Inc. (2025). Gartner hype cycle identifies top ai innovations in 2025. <https://www.gartner.com/en/newsroom/press-releases/2025-08-05-gartner-hype-cycle-identifies-top-ai-innovations-in-2025>. Accessed: 2025-09-21. 5.2.2
- Géron, Aurélien (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media. 5.1.4
- Ministerio de Energía (2021). Decreto 67: Establece Especificaciones de Calidad para Gasolina de Aviación. <https://bcn.cl/2uled>. Accessed: 2025-09-04. 4.2.2
- Nieto, Bely Iván Melgar (2022). Machine learning en la predicción de propiedades pvt. Director de Tesis: Dr. Víctor Leonardo Teja Juárez. 4.2.5
- Osterwalder, Alexander y Pigneur, Yves (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, NJ: John Wiley & Sons. 5.2.3
- Zoting, Shivani (2025). Machine learning market size, share & trends 2025 to 2034. <https://www.precedenceresearch.com/machine-learning-market>. Accessed: 2025-09-21. 5.2.2

A | Especificación de Calidad de Gasolina de Aviación

Requisitos (i)	Límite	Unidad	Valor	Métodos de ensayo ASTM
Densidad a 15 °C	-	kg/m ³	Informar	D4052, D1298
Apariencia (ii)	-		Informar	
Antioxidante	Máximo	g/m ³	12	-
Azufre	Máximo	% m/m	0,05	D2622, D1266, D5453
Calor neto de combustión	Mínimo	MJ/kg	43,5	D4529, D3338
Color	-	-	Azul	D2392
Contenido de colorante (iii):				
- azul	Máximo	mg/l	2,7	-
- amarillo	Máximo	mg/l	ninguno	-
Corrosión de la lámina de cobre 2 Horas a 100 C°	Máximo	Nº	1	D130
Destilación:				
- Punto inicial	-	°C	Informar	
- 10 % evaporado	Máximo	°C	75	
- 40 % evaporado	Mínimo	°C	75	
- 50 % evaporado	Máximo	°C	105	
- 90 % evaporado	Máximo	°C	135	
- Punto final	Máximo	°C	170	D86
- Suma de las temperaturas al 10% evaporado y al 50% evaporado	Mínimo	°C	135	
- Recuperado	Mínimo	% v/v	97	
- Residuo de destilación	Máximo	% v/v	1,5	
- Pérdida por destilación	Máximo	% v/v	1,5	
Detonancia mezcla pobre :		-		
- NOM (Número Octano Motor)	Mínimo	Nº	99,6	D2700
- Graduación pobre de aviación	Mínimo	Nº	100,0	
Detonancia mezcla rica,				
- Número de comportamiento (iv)	Mínimo	Nº	130,0	D909
Estabilidad a la oxidación (5 hrs.):				
- Goma potencial (v)	Máximo	mg/100 ml	6	D873
- Precipitado de plomo	Máximo	mg/100 ml	3	
Presión de vapor	Mínimo	kPa	38	
	Máximo	kPa	49	D323, D5191(vi)
Punto de congelación	Máximo	°C	- 58 (vii)	D2386
Reacción al agua, variación de volumen	Máximo	ml	± 2	D1094
Tetraetilo de plomo (viii)	Mínimo	g Pb/l (ml TEL/l)	0,28 (0,27)	
	Máximo	g Pb/l (ml TEL/l)	0,56 (0,53)	D3341, D5059
Conductividad eléctrica	Máximo	pS/m	450 (ix)	D2624

Figura A.1: Requisitos y especificaciones nacionales de calidad para gasolina de aviación

(Fuente: Ministerio de Energía, Biblioteca del Congreso Nacional)

B | Código Fuente Notebook Jupyter

```

1 # COMENZAMOS EL ESTUDIO DE LOS DATOS PARA DESARROLLAR EL MODELO
2 # IMPORTAMOS LIBRERÍAS BÁSICAS
3 import pandas as pd
4 import numpy as np
5 import matplotlib.pyplot as plt
6 from sklearn.preprocessing import LabelEncoder
7 from sklearn.preprocessing import OneHotEncoder
8 from sklearn.compose import ColumnTransformer
9
10 # Lo que vamos a cargar para revisar es una tabla prefiltrada
11 # esta se obtuvo a partir de los datos originales de las tablas de winblis y
    samplemanager
12 # como no tenemos acceso directo a la base de datos del lims, se debe hacer
    manualmente
13 # este archivo de excel contiene solo las columnas que nos interesan
14 # además, se han eliminado las filas con datos faltantes
15 # también se han eliminado las columnas que no nos interesan
16 # el archivo se llama "gaso_aviac_v2.xlsx"
17
18 df = pd.read_excel("GASO_AVIAC_v2.xlsx")
19 df

```

	Numero de Secuencia	Punto de Muestreo	Densidad a 15 °C	kg/m3	\
0	128640	416B		698.9	
1	135822	416A		698.6	
2	137291	416B		698.6	
3	146867	416B		698.6	
4	149686	416A		698.6	
..	
253	1648803	416B		697.7	
254	1672648	416B		696.8	
255	1696619	416B		696.8	
256	1726331	416B		696.9	
257	1750488	416B		696.7	

	API Gravedad, Digital	60°F	Punto inicial °C	10 % °C	20 % °C	30 % °C	\
0		70.9	40.0	74.0	86.0	94.0	
1		71.0	38.0	72.0	85.0	94.0	
2		71.0	39.0	70.0	84.0	93.0	
3		71.0	39.0	74.0	86.0	95.0	
4		71.0	40.0	74.0	86.0	94.0	
..		
253		71.3	32.0	71.3	88.5	97.1	
254		71.4	30.4	69.3	86.5	96.1	
255		71.5	30.3	69.5	85.8	95.8	
256		71.5	32.6	69.0	85.3	95.2	
257		71.6	32.0	68.5	84.9	94.9	

	40 % °C	50 % °C	...	90 % °C	95 % °C	Punto final °C	\
0	100.0	102.0	...	114.0	120.0	160.0	
1	100.0	103.0	...	114.0	120.0	162.0	
2	99.0	103.0	...	114.0	124.0	162.0	
3	100.0	104.0	...	114.0	124.0	164.0	
4	98.0	102.0	...	114.0	120.0	158.0	
..	
253	101.2	103.7	...	112.1	116.8	133.2	
254	100.8	103.2	...	111.7	115.8	131.1	
255	100.8	103.6	...	112.2	116.7	132.5	

```

37 256    100.4    103.2    ...    112.2    117.0                135.8
38 257    100.1    102.8    ...    111.7    116.1                130.4
39
40      Recuperado % v/v  Residuo % v/v  Pérdida % v/v  Colorante azul g/m3  \
41 0          98.0        1.0            1.0            1.020
42 1          98.0        1.2            0.8            1.290
43 2          98.0        1.0            1.0            1.150
44 3          98.0        1.0            1.0            1.390
45 4          98.0        1.0            1.0            1.200
46 ..          ...        ...            ...            ...
47 253        98.1        0.9            1.0            2.089
48 254        98.3        0.9            0.8            2.200
49 255        98.2        0.9            0.9            2.310
50 256        97.7        0.9            1.4            2.040
51 257        97.7        0.9            1.4            2.050
52
53      Presión de vapor psi  Plomo g/L  Detonancia mezcla rica
54 0          43.40        0.71            135.0
55 1          43.90        0.50            132.0
56 2          44.20        0.44            131.0
57 3          45.20        0.61            133.0
58 4          43.40        0.51            132.0
59 ..          ...        ...            ...
60 253        6.52        0.46            133.0
61 254        6.82        0.46            133.1
62 255        6.88        0.52            134.6
63 256        6.53        0.48            133.8
64 257        6.61        0.49            134.1
65
66 [258 rows x 23 columns]

```

```

1 # Revisar el tipo de datos de cada columna
2
3 df.dtypes.value_counts()

```

```

1 float64    21
2 int64      1
3 object     1
4 Name: count, dtype: int64

```

```

1 # Valores Nulos
2 # verificamos si hay columnas que tienen aún valores nulos
3 # la idea es que tomemos solo las columnas que no tienen valores nulos
4 # o eliminar las filas que los contengan
5 # ejecutamos este bloque limpiando hasta que todas las columnas tengan 0
6   valores nulos
7
7 for feature in df.columns:
8     print('Total de valores nulos de', feature, '=', df[feature].isna().sum())

```

```

1 Total de valores nulos de Numero de Secuencia = 0
2 Total de valores nulos de Punto de Muestreo = 0
3 Total de valores nulos de Densidad a 15 °C kg/m3 = 0
4 Total de valores nulos de API Gravedad, Digital 60°F = 0
5 Total de valores nulos de Punto inicial °C = 0
6 Total de valores nulos de 10 % °C = 0
7 Total de valores nulos de 20 % °C = 0

```

```

8 Total de valores nulos de 30 % °C = 0
9 Total de valores nulos de 40 % °C = 0
10 Total de valores nulos de 50 % °C = 0
11 Total de valores nulos de 60 % °C = 0
12 Total de valores nulos de 70 % °C = 0
13 Total de valores nulos de 80 % °C = 0
14 Total de valores nulos de 90 % °C = 0
15 Total de valores nulos de 95 % °C = 0
16 Total de valores nulos de Punto final °C = 0
17 Total de valores nulos de Recuperado % v/v = 0
18 Total de valores nulos de Residuo % v/v = 0
19 Total de valores nulos de Pérdida % v/v = 0
20 Total de valores nulos de Colorante azul g/m3 = 0
21 Total de valores nulos de Presión de vapor psi = 0
22 Total de valores nulos de Plomo g/L = 0
23 Total de valores nulos de Detonancia mezcla rica = 0

```

```

1 # FORMA # 23 Columnas; 258 Filas #
2
3 df.shape

```

```

1 (258, 23)

```

```

1 # Limpieza de columnas nulas
2 # Hacer esto si se requiere limpiar más el dataframe
3 # En el último caso el Excel viene limpio
4 # Elimino también las columnas que no aportan o interesan en el estudio
5
6 df_limpio = df
7
8 df_limpio = df_limpio.drop('Numero_de_Secuencia', axis=1)
9 df_limpio = df_limpio.drop('Punto_de_Muestreo', axis=1)
10 #df_limpio = df_limpio.drop('Corrosión Lámina de cobre (2 h a 100°C)', axis=1)
11 #df_limpio = df_limpio.drop('Olefinas', axis=1)
12 #df_limpio = df_limpio.drop('Inferior, API Gravedad, Digital 60°F.1', axis=1)
13 #df_limpio = df_limpio.drop('Media, API Gravedad, Digital 60°F.1', axis=1)
14 #df_limpio = df_limpio.drop('Superior, API Gravedad, Digital 60°F', axis=1)
15 #df_limpio = df_limpio.drop('Inferior, API Gravedad, Digital 60°F', axis=1)
16 #df_limpio = df_limpio.drop('Media, API Gravedad, Digital 60°F', axis=1)
17
18 df_limpio

```

	Densidad a 15 °C	kg/m3	API Gravedad, Digital	60°F	Punto inicial °C	\		
0		698.9		70.9	40.0			
1		698.6		71.0	38.0			
2		698.6		71.0	39.0			
3		698.6		71.0	39.0			
4		698.6		71.0	40.0			
..				
253		697.7		71.3	32.0			
254		696.8		71.4	30.4			
255		696.8		71.5	30.3			
256		696.9		71.5	32.6			
257		696.7		71.6	32.0			
	10 % °C	20 % °C	30 % °C	40 % °C	50 % °C	60 % °C	70 % °C	... \
0	74.0	86.0	94.0	100.0	102.0	105.0	107.0	...

```

16 1      72.0    85.0    94.0    100.0    103.0    105.0    108.0    ...
17 2      70.0    84.0    93.0    99.0     103.0    105.0    108.0    ...
18 3      74.0    86.0    95.0    100.0    104.0    106.0    108.0    ...
19 4      74.0    86.0    94.0    98.0     102.0    104.0    106.0    ...
20 ..     ...     ...     ...     ...     ...     ...     ...     ...
21 253    71.3     88.5     97.1    101.2    103.7    105.4    107.0    ...
22 254    69.3     86.5     96.1    100.8    103.2    105.0    106.6    ...
23 255    69.5     85.8     95.8    100.8    103.6    105.5    107.1    ...
24 256    69.0     85.3     95.2    100.4    103.2    105.2    106.9    ...
25 257    68.5     84.9     94.9    100.1    102.8    104.7    106.5    ...
26
27      90 % °C  95 % °C  Punto final °C  Recuperado % v/v  Residuo % v/v  \
28 0      114.0    120.0    160.0           98.0           1.0
29 1      114.0    120.0    162.0           98.0           1.2
30 2      114.0    124.0    162.0           98.0           1.0
31 3      114.0    124.0    164.0           98.0           1.0
32 4      114.0    120.0    158.0           98.0           1.0
33 ..     ...     ...     ...           ...           ...
34 253    112.1    116.8    133.2           98.1           0.9
35 254    111.7    115.8    131.1           98.3           0.9
36 255    112.2    116.7    132.5           98.2           0.9
37 256    112.2    117.0    135.8           97.7           0.9
38 257    111.7    116.1    130.4           97.7           0.9
39
40      Pérdida % v/v  Colorante azul g/m3  Presión de vapor psi  Plomo g/L  \
41 0      1.0           1.020           43.40           0.71
42 1      0.8           1.290           43.90           0.50
43 2      1.0           1.150           44.20           0.44
44 3      1.0           1.390           45.20           0.61
45 4      1.0           1.200           43.40           0.51
46 ..     ...           ...           ...           ...
47 253    1.0           2.089           6.52           0.46
48 254    0.8           2.200           6.82           0.46
49 255    0.9           2.310           6.88           0.52
50 256    1.4           2.040           6.53           0.48
51 257    1.4           2.050           6.61           0.49
52
53      Detonancia mezcla rica
54 0      135.0
55 1      132.0
56 2      131.0
57 3      133.0
58 4      132.0
59 ..     ...
60 253    133.0
61 254    133.1
62 255    134.6
63 256    133.8
64 257    134.1
65
66 [258 rows x 21 columns]

```

```

1 # REVISION DE FORMA
2
3 df_limpio.shape

```

```

1 (258, 21)

```

```

1 # NULOS POR FILA # BUSCANDO INDICE #
2
3 filas_con_nulos = df_limpio[df_limpio.isnull().any(axis=1)]
4 conteo_nulos_por_fila = filas_con_nulos.isnull().sum(axis=1)
5
6 info_filas_nulos = pd.DataFrame({
7     'Índice': filas_con_nulos.index,
8     'N° MUESTRA': filas_con_nulos.iloc[:, 0],
9     'Cantidad_de_Nulos': conteo_nulos_por_fila
10 })
11
12 print(info_filas_nulos)

```

```

1 Empty DataFrame
2 Columns: [Índice, N° MUESTRA, Cantidad de Nulos]
3 Index: []

```

```

1 # ELIMINANDO FILAS CON NULOS
2 # solo ejecutar cuando se estén limpiando los datos
3
4 #df_limpio = df_limpio.drop(9)

```

```

1 # REVISION DE FORMA
2
3 df_limpio.shape

```

```

1 (258, 21)

```

```

1 # Gráfico de Correlación (Calor)
2
3 import seaborn as sns
4
5 columna_foco = 'Detonancia_mezcla_rica' # Columna Target
6 columnas_ordenadas = df_limpio.columns.tolist()
7 columnas_ordenadas.remove(columna_foco)
8 columnas_ordenadas = columnas_ordenadas + [columna_foco] #Asignar la columna
9     foco al final del dataset
10 df_limpio = df_limpio[columnas_ordenadas]
11
12 matriz_correlacion = df_limpio.corr()
13
14 plt.figure(figsize=(24, 16))
15 sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm', fmt=".2f",
16     linewidths=.5)
17 plt.title('')
18 plt.show()

```

```

1 <Figure size 2400x1600 with 2 Axes>

```

```

1 #EVALUANDO EL GRAFICO DE CORRELACIÓN
2
3 # Aquí hay que verificar si hay alguna categoría que no generen correlación
4 # Dejaré comentado al menos 7 categorías que se obtuvieron con datos sin
5     cambios en el proceso iterativo de limpieza de datos
6 #

```

```

6 # Eliminación de columnas sin variabilidad de los datos
7
8 #df_limpio = df_limpio.drop('Aspecto de Interfase', axis=1)
9 #df_limpio = df_limpio.drop('Azufre', axis=1)
10 #df_limpio = df_limpio.drop('Azufre [ppm]', axis=1)
11 #df_limpio = df_limpio.drop('Color', axis=1)
12 #df_limpio = df_limpio.drop('Mono-Aromáticos', axis=1)
13 #df_limpio = df_limpio.drop('Poli-Aromáticos', axis=1)
14 #df_limpio = df_limpio.drop('TOTAL', axis=1)

```

```

1 # REVISION DE FORMA # 37 Columnas; 23 Filas #
2
3 df_limpio.shape

```

```

1 (258, 21)

```

```

1 # MAPA DE CALOR FINAL #
2
3 matriz_correlacion = df_limpio.corr()
4
5 plt.figure(figsize=(24, 16))
6 sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm', fmt=".2f",
7             linewidths=.5)
8 plt.title('MAPA_DE_CALOR')
9 plt.show()

```

```

1 <Figure size 2400x1600 with 2 Axes>

```

```

1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.linear_model import LinearRegression
4 import matplotlib.pyplot as plt
5
6 # Dividir el DataFrame en características (X) y variable objetivo (y)
7 X = df_limpio.drop('Detonancia_mezcla_rica', axis=1)
8 y = df_limpio['Detonancia_mezcla_rica']
9
10 # Dividir el conjunto de datos en conjuntos de entrenamiento y prueba
11 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
12                                                    random_state=42)
13
14 # Crear y entrenar el modelo de regresión lineal
15 model = LinearRegression()
16 model.fit(X_train, y_train)
17
18 # Hacer predicciones en el conjunto de prueba
19 y_pred = model.predict(X_test)

```

```

1 df_sin_target = df_limpio.drop('Detonancia_mezcla_rica', axis=1)
2
3 df_sin_target

```

```

1      Densidad a 15 °C kg/m3  API Gravedad, Digital 60°F  Punto inicial °C \
2  0          698.9          70.9          40.0
3  1          698.6          71.0          38.0
4  2          698.6          71.0          39.0

```

5	3			698.6			71.0		39.0
6	4			698.6			71.0		40.0
7
8	253			697.7			71.3		32.0
9	254			696.8			71.4		30.4
10	255			696.8			71.5		30.3
11	256			696.9			71.5		32.6
12	257			696.7			71.6		32.0
13									
14		10 % °C	20 % °C	30 % °C	40 % °C	50 % °C	60 % °C	70 % °C	80 % °C \
15	0	74.0	86.0	94.0	100.0	102.0	105.0	107.0	110.0
16	1	72.0	85.0	94.0	100.0	103.0	105.0	108.0	110.0
17	2	70.0	84.0	93.0	99.0	103.0	105.0	108.0	110.0
18	3	74.0	86.0	95.0	100.0	104.0	106.0	108.0	110.0
19	4	74.0	86.0	94.0	98.0	102.0	104.0	106.0	110.0
20
21	253	71.3	88.5	97.1	101.2	103.7	105.4	107.0	109.0
22	254	69.3	86.5	96.1	100.8	103.2	105.0	106.6	108.5
23	255	69.5	85.8	95.8	100.8	103.6	105.5	107.1	109.0
24	256	69.0	85.3	95.2	100.4	103.2	105.2	106.9	108.9
25	257	68.5	84.9	94.9	100.1	102.8	104.7	106.5	108.3
26									
27		90 % °C	95 % °C	Punto final °C	Recuperado % v/v	Residuo % v/v			\
28	0	114.0	120.0	160.0	98.0	1.0			
29	1	114.0	120.0	162.0	98.0	1.2			
30	2	114.0	124.0	162.0	98.0	1.0			
31	3	114.0	124.0	164.0	98.0	1.0			
32	4	114.0	120.0	158.0	98.0	1.0			
33			
34	253	112.1	116.8	133.2	98.1	0.9			
35	254	111.7	115.8	131.1	98.3	0.9			
36	255	112.2	116.7	132.5	98.2	0.9			
37	256	112.2	117.0	135.8	97.7	0.9			
38	257	111.7	116.1	130.4	97.7	0.9			
39									
40		Pérdida % v/v	Colorante azul g/m3	Presión de vapor psi	Plomo g/L				
41	0	1.0	1.020	43.40	0.71				
42	1	0.8	1.290	43.90	0.50				
43	2	1.0	1.150	44.20	0.44				
44	3	1.0	1.390	45.20	0.61				
45	4	1.0	1.200	43.40	0.51				
46				
47	253	1.0	2.089	6.52	0.46				
48	254	0.8	2.200	6.82	0.46				
49	255	0.9	2.310	6.88	0.52				
50	256	1.4	2.040	6.53	0.48				
51	257	1.4	2.050	6.61	0.49				
52									
53									

[258 rows x 20 columns]

```

1 df_sin_target['y_pred'] = model.predict(df_sin_target)
2
3 df_sin_target['y_Real'] = y
4
5 df_sin_target

```

1		Densidad a 15 °C kg/m3	API Gravedad, Digital 60°F	Punto inicial °C \
2	0	698.9	70.9	40.0

3	1					698.6		71.0		38.0
4	2					698.6		71.0		39.0
5	3					698.6		71.0		39.0
6	4					698.6		71.0		40.0
7
8	253					697.7		71.3		32.0
9	254					696.8		71.4		30.4
10	255					696.8		71.5		30.3
11	256					696.9		71.5		32.6
12	257					696.7		71.6		32.0
13										
14		10 % °C	20 % °C	30 % °C	40 % °C	50 % °C	60 % °C	70 % °C	...	\
15	0	74.0	86.0	94.0	100.0	102.0	105.0	107.0	...	
16	1	72.0	85.0	94.0	100.0	103.0	105.0	108.0	...	
17	2	70.0	84.0	93.0	99.0	103.0	105.0	108.0	...	
18	3	74.0	86.0	95.0	100.0	104.0	106.0	108.0	...	
19	4	74.0	86.0	94.0	98.0	102.0	104.0	106.0	...	
20	
21	253	71.3	88.5	97.1	101.2	103.7	105.4	107.0	...	
22	254	69.3	86.5	96.1	100.8	103.2	105.0	106.6	...	
23	255	69.5	85.8	95.8	100.8	103.6	105.5	107.1	...	
24	256	69.0	85.3	95.2	100.4	103.2	105.2	106.9	...	
25	257	68.5	84.9	94.9	100.1	102.8	104.7	106.5	...	
26										
27		95 % °C	Punto final °C	Recuperado	% v/v	Residuo	% v/v	Pérdida	% v/v	\
28	0	120.0	160.0		98.0		1.0		1.0	
29	1	120.0	162.0		98.0		1.2		0.8	
30	2	124.0	162.0		98.0		1.0		1.0	
31	3	124.0	164.0		98.0		1.0		1.0	
32	4	120.0	158.0		98.0		1.0		1.0	
33	
34	253	116.8	133.2		98.1		0.9		1.0	
35	254	115.8	131.1		98.3		0.9		0.8	
36	255	116.7	132.5		98.2		0.9		0.9	
37	256	117.0	135.8		97.7		0.9		1.4	
38	257	116.1	130.4		97.7		0.9		1.4	
39										
40		Colorante azul g/m3	Presión de vapor psi	Plomo g/L	y_pred	y_Real				
41	0	1.020		43.40	135.547300	135.0				
42	1	1.290		43.90	133.155645	132.0				
43	2	1.150		44.20	131.388288	131.0				
44	3	1.390		45.20	134.662873	133.0				
45	4	1.200		43.40	132.171629	132.0				
46				
47	253	2.089		6.52	133.021736	133.0				
48	254	2.200		6.82	132.810790	133.1				
49	255	2.310		6.88	133.975121	134.6				
50	256	2.040		6.53	132.601932	133.8				
51	257	2.050		6.61	132.472331	134.1				
52										
53										

[258 rows x 22 columns]

```

1 from sklearn.linear_model import LinearRegression, ElasticNet
2 from sklearn.ensemble import RandomForestRegressor
3 from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error
4 from sklearn.model_selection import train_test_split
5 import xgboost as xgb
6
7 # Ahora puedes dividir los datos en conjunto de entrenamiento y prueba

```

```

8 from sklearn.model_selection import train_test_split
9 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
10 random_state=42)
11 # Dividir el conjunto de datos en conjuntos de entrenamiento y prueba
12 #X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
13 random_state=42)
14 # Función para evaluar un modelo y calcular MSE y Error Promedio
15 def evaluar_modelo(modelo, X_train, X_test, y_train, y_test):
16     modelo.fit(X_train, y_train)
17     y_pred = modelo.predict(X_test)
18     mse = mean_squared_error(y_test, y_pred)
19     error_promedio = mean_absolute_percentage_error(y_test, y_pred) * 100 #
20     Error como porcentaje
21     return mse, error_promedio
22 # 1. Regresión Lineal
23 modelo_lr = LinearRegression()
24 mse_lr, error_lr = evaluar_modelo(modelo_lr, X_train, X_test, y_train, y_test)
25
26 # 2. Random Forest Regressor
27 modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)
28 mse_rf, error_rf = evaluar_modelo(modelo_rf, X_train, X_test, y_train, y_test)
29
30 # 3. Elastic Net
31 modelo_en = ElasticNet(random_state=42)
32 mse_en, error_en = evaluar_modelo(modelo_en, X_train, X_test, y_train, y_test)
33
34 # 4. XGBoost
35 modelo_xgb = xgb.XGBRegressor(random_state=42)
36 mse_xgb, error_xgb = evaluar_modelo(modelo_xgb, X_train, X_test, y_train,
37 y_test)
38 # Mostrar resultados
39 print(f"Resultado_Regresion_Lineal\nMSE_promedio: {mse_lr:.3f}\nError_promedio:
40 {error_lr:.2f}%\n")
41 print(f"Resultado_Random_Forest_Regressor\nMSE_promedio: {mse_rf:.3f}\nError_
42 promedio: {error_rf:.2f}%\n")
43 print(f"Resultado_Elastic_Net\nMSE_promedio: {mse_en:.3f}\nError_promedio: {
44 error_en:.2f}%\n")
45 print(f"Resultado_XGBoost\nMSE_promedio: {mse_xgb:.3f}\nError_promedio: {
46 error_xgb:.2f}%\n")

```

```

1 Resultado Regresion Lineal
2 MSE promedio: 2.888
3 Error promedio: 1.01%
4
5 Resultado Random Forest Regressor
6 MSE promedio: 2.567
7 Error promedio: 0.93%
8
9 Resultado Elastic Net
10 MSE promedio: 2.765
11 Error promedio: 0.99%
12
13 Resultado XGBoost
14 MSE promedio: 2.851
15 Error promedio: 0.97%

```

```

1 from sklearn.model_selection import train_test_split
2 from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error
3
4 # Asegúrate de que X e y estén correctamente definidos (ya lo has hecho en los
5   pasos anteriores)
6 # X = df_limpio.drop('Detonancia mezcla rica', axis=1)
7 # y = df_limpio['Detonancia mezcla rica']
8
9 # Dividir los datos en conjunto de entrenamiento y prueba
10 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
11   random_state=42)
12
13 # Crear el modelo RandomForest
14 modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)
15
16 # Entrenar el modelo
17 modelo_rf.fit(X_train, y_train)
18
19 # Hacer predicciones en el conjunto de prueba
20 y_pred = modelo_rf.predict(X_test)
21
22 # Calcular el MSE y el error promedio
23 mse_rf = mean_squared_error(y_test, y_pred)
24 error_rf = mean_absolute_percentage_error(y_test, y_pred) * 100
25
26 # Mostrar los resultados
27 print(f"Resultado Random Forest\nMSE promedio: {mse_rf:.3f}\nError promedio: {
28   error_rf:.2f}%\n")
29
30 y_pred
31 modelo_rf
32
33 import joblib
34
35 # Serializar el modelo Random Forest en un archivo .pkl
36 archivo_modelo = "modelo_rf.pkl"
37 joblib.dump(modelo_rf, archivo_modelo)
38
39 print(f"El modelo Random Forest ha sido guardado en: {archivo_modelo}")

```

```

1 Resultado Random Forest
2 MSE promedio: 2.250
3 Error promedio: 0.81%
4
5 El modelo Random Forest ha sido guardado en: modelo_rf.pkl

```

```

1 #Visualizar el ultimo modelo entrenado con muestra de datos
2 #Valor real vs predicho
3
4 X_test["Detonancia_mezcla_rica"] = y_test
5 X_test["Predicción"] = y_pred
6
7 X_test[["Detonancia_mezcla_rica", "Predicción"]].head(20)

```

```

1 Detonancia mezcla rica Predicción
2 66 132.0 133.336

```

3	45	134.0	134.910
4	9	132.0	133.200
5	73	134.0	133.502
6	136	132.0	134.987
7	104	130.0	132.738
8	205	134.0	133.496
9	247	130.9	132.225
10	125	136.0	134.532
11	211	132.5	132.475
12	84	132.0	133.481
13	33	134.0	134.070
14	254	133.1	132.866
15	68	136.0	134.021
16	46	136.0	134.375
17	108	131.0	131.710
18	6	133.0	133.069
19	30	133.0	132.945
20	101	134.0	133.251
21	55	134.0	132.132