

DE MODELOS PRESCRIPTIVOS A PROACTIVOS – DATA SCIENCE: UNA OPORTUNIDAD PARA LA GESTION DE INTEGRIDAD

Diego Sebastián López, Consultor, diego.lopez1991@gmail.com

Sinopsis

Lo marcos normativos y las prácticas recomendadas de gestión de integridad suelen definir dos posibles modelos de gestión de integridad.

Los modelos prescriptivos establecen metodologías y plazos preestablecidos para realizar las evaluaciones de integridad. Además, están integrados dentro de un proceso de mejora continua que busca capitalizar las lecciones aprendidas y optimizar el proceso.

Por otro lado, también contemplan la opción de migrar desde un modelo prescriptivo a uno proactivo, donde las definiciones y plazos no estén preestablecidos, sino que se determinan según la condición del activo. De esa manera, la operación puede ganar eficiencia, competitividad y generar optimizar los costos operativos mientras se mitigan los riesgos de manera dirigida.

Para ello, es necesario capturar valor del comportamiento previo de los activos y desarrollar modelos confiables. El presente trabajo describe el desarrollo y validación de un modelo predictivo de integridad utilizando la base de datos AIIM Onshore Pipeline Database [1], la cual integra más de 350 variables ambientales y operativas. Además, se desarrolla un sistema para priorizar intervenciones en función del riesgo y se estiman los beneficios económicos asociados a la implementación de un sistema de integridad proactivo basado en ciencia de datos.

La metodología propuesta se aleja de los calendarios fijos de inspección para centrarse en la condición y el entorno real del activo. Mediante técnicas de ciencia de datos (Data Science), se realizó un preprocesamiento exhaustivo y se compararon modelos estadísticos clásicos, como la Regresión Logística, con algoritmos de aprendizaje automático, específicamente Random Forest.

Los resultados demuestran que el modelo de Random Forest classifier (binario falla/no falla) presenta una capacidad superior para identificar segmentos con mayor susceptibilidad a fallas por corrosión, logrando un área bajo la curva (AUC) significativamente mayor que los métodos tradicionales, lo cual implica que el modelo es altamente eficaz para predecir condición.

Adicionalmente al análisis del rendimiento técnico de los modelos, el trabajo introduce un análisis de captura de riesgo económico. Se demuestra que, mediante la jerarquización predictiva, es posible capturar una fracción dominante del riesgo económico (representado por el valor de construcción y agrícola de las áreas afectadas) asignando recursos en solo una pequeña fracción de los activos. Particularmente, limitando las actividades a un determinado porcentaje de los activos (ej: 20%), el modelo permitió capturar un 40% más de riesgo económico esperado respecto a modelos estadísticos clásicos (92,6% vs 65,3), mejorando sustancialmente la eficiencia en la asignación de recursos bajo restricciones presupuestarias.

Introducción

1.1 Motivación

La gestión de integridad de ductos y activos físicos es un desafío estratégico que combina aspectos técnicos, regulatorios y económicos que dependen del entorno de operación.

En algunos casos, las características de los activos generan alta probabilidad de falla y bajas consecuencias; en otros casos generan baja probabilidad de falla y alta consecuencia; o bien, se pueden presentar casos intermedios.

En este contexto, **las normas y prácticas recomendadas han establecido históricamente modelos prescriptivos, donde las inspecciones y evaluaciones se realizan bajo metodologías y plazos predefinidos.** Este enfoque, si bien aporta orden y estandarización, **puede resultar limitado frente a la creciente complejidad de los activos y la necesidad de optimizar recursos.**

Además, actualmente, **las compañías enfrentan la necesidad de maximizar la confiabilidad, cumplir con regulaciones y simultáneamente reducir costos. Por eso, es necesario establecer un lenguaje común entre las áreas técnicas y las áreas económicas** que ofrezca una traducción clara entre ambos mundos, evitando así que las actividades de integridad pueden percibirse como gastos en lugar de inversiones estratégicas que aportan valor.

Una posible solución para establecer un lenguaje común **consiste en recurrir a la incorporación de nuevas tecnologías y modelos que permiten obtener predicciones confiables sobre el comportamiento de un activo y cuantificar beneficios** asociados a estas predicciones.

En los últimos años, **la ciencia de datos y la inteligencia artificial han demostrado un potencial significativo para abordar este desafío.** Modelos estadísticos clásicos, como regresiones y análisis de series temporales, ofrecen un punto de partida para interpretar sesgos y limitaciones de la información. A partir de allí, técnicas avanzadas como Random Forest, Gradient Boosting, redes neuronales y redes bayesianas permiten identificar relaciones complejas entre variables y construir sistemas predictivos de integridad. Estas metodologías replican fenómenos naturales o procesos cognitivos humanos, generando modelos robustos capaces de integrar grandes volúmenes de datos.

El presente trabajo se enmarca en esta tendencia y propone una visión estratégica para la gestión de integridad, utilizando como caso de estudio la AIIM (Advanced Infrastructure Integrity Modeling) Onshore Pipeline Database [1] desarrollada por el NETL (National Energy Technology Laboratory) del DOE (U.S. Department of Energy). Este dataset, de acceso abierto y con más de 350 variables ambientales y operativas, constituye una base sólida para explorar la aplicación de modelos de ciencia de datos en la transición de un enfoque prescriptivo hacia uno proactivo. A partir de su análisis, se busca demostrar cómo la digitalización y el uso de inteligencia artificial pueden generar beneficios cuantitativos y cualitativos para la industria.

1.2 Estado del arte en integridad de activos

La gestión de integridad de activos en la industria del petróleo y gas ha evolucionado significativamente en las últimas décadas, impulsada por la necesidad de reducir riesgos y cumplir con regulaciones cada vez más estrictas. Este proceso se ha visto reflejado en la transición desde modelos prescriptivos hacia enfoques apoyados fundamentalmente en metodologías de análisis de riesgo o bien, metodologías/prácticas recomendadas específicas de gestión de integridad que abordan alguna amenaza en particular.

Las metodologías más avanzadas son habitualmente desarrolladas por organizaciones profesionales como pueden ser AMPP (The Association for Materials Protection and Performance), ASME (The American Society of Mechanical Engineers), API (The American Petroleum Institute), entre otras.

A modo de ejemplo, se citan las metodologías para evaluación de corrosión externa [2] e interna [3] que son referencia global en gestión de corrosión de ductos y sistemas de transporte, permitiendo identificar zonas críticas y tomar decisiones de revaluación. Estas metodologías suelen implementarse dentro de programas de gestión de integridad tales como [4] y [5], que ofrecen un sistema estándar para gestión de integridad, incluyendo identificación de amenazas, análisis de riesgo y evaluaciones de integridad dentro de un proceso de mejora continua.

En cuanto a los **modelos de riesgo para ductos**, se han desarrollado enfoques que se dividen en modelos semicuantitativos y cuantitativos. Entre los primeros, **el modelo de Muhlbauer**, difundido a través del Pipeline Risk Management Manual [6], se ha convertido en una referencia práctica para operadores de ductos. Este enfoque **utiliza un sistema de índices ponderados para estimar un puntaje de riesgo relativo** por segmento de ducto y permite obtener un ranking, facilitando la priorización de inspecciones y actividades de mantenimiento o gestión de activos. Su principal fortaleza es la simplicidad y la aplicabilidad rápida, **aunque presenta limitaciones asociadas a la subjetividad en la asignación de puntajes y a la dificultad de capturar relaciones complejas entre variables**. En el ámbito de los **modelos cuantitativos**, normas como la BS 8010 parte 3 [7] y la IGEM TD/1 [8] **han aportado mayor rigor técnico, aunque suelen utilizar datos históricos en lugar de modelado estadístico predictivo**. El desarrollo de bases de datos ha sido clave para alimentar estos modelos: la PHMSA [9] en Estados Unidos, el UKOPA [10] en Reino Unido, entre otros han generado repositorios de incidentes que sirven como fuentes de información para los análisis. En Argentina, existe el reporte RIDA [11] publicado por el IAPG.

Para el caso particular de equipos, se cita la metodología de gestión de riesgos según las prácticas recomendadas que desarrollan los elementos y la metodología para **programas de inspección basada en riesgo** (Risk Based Inspection – RBI [12] y [13]). Estas prácticas permiten desarrollar programas que **combinan la probabilidad de falla con las consecuencias potenciales** de manera similar a los análisis de riesgo cuantitativos, sustentando la toma de decisiones y la priorización de recursos en los activos más críticos, y superando la rigidez de los calendarios prescriptivos.

Por otro lado, **las normas ISO 55000 [14], ISO 55001 [15] e ISO 55002 [16] ofrecen** una visión más amplia, moderna y vanguardista, proponiendo **un enfoque integral que vincula la gestión técnica con la estrategia corporativa y los objetivos económicos**. Refuerzan la idea de que la integridad no es solo un tema técnico, sino también financiero y organizacional.

Desarrollo

2.1 Dataset AIIM: descripción y limitaciones

El caso de estudio se basa en la AIIM Onshore Pipeline Database [1], desarrollada por el NETL del DOE. Se trata de un dataset abierto, que contiene información detallada sobre ductos terrestres (onshore) en EE.UU. y Alaska.

- Cobertura: **más de 500.000 puntos** con información geográfica.
- Variables: **más de 350 atributos** que incluyen datos operativos como el tipo de fluido, la cantidad de incidentes reportados y clasificados, características y propiedades del suelo, condiciones del entorno tales como presencia de cuerpos de agua, distancia a

interferencias, uso de suelo, valor económico de la construcción y valor económico agrícola, entre otros.

- Interpretación de variables: **representan las características, propiedades y/o antecedentes del punto geográfico y su entorno, es decir, son variables representativas y de agregación.** Por ejemplo, para las variables de recuento de fallas, se suman las ocurrencias en todos los segmentos de ductos ubicado dentro del área de agregación. Esto permite que los contadores reflejen eventos de integridad ocurridos en la vecindad de cada segmento, y no valores globales promediados por toda la longitud de la cañería.
- Formato: capas GIS (archivo .gdb)
- **Limitaciones: no incluye la antigüedad de los ductos** ni detalla la serie temporal de las fallas. Estas ausencias restringen el análisis directo de las amenazas dependientes del tiempo.

2.2 Metodología de ciencia de datos

El análisis se implementó en el **entorno Anaconda**, utilizando **lenguaje Python** y se estructuró en ocho etapas principales que permitieron optimizar la cantidad de variables a incorporar en los modelos predictivos, mediante conceptos estadísticos, técnicos y analíticos. A continuación se desarrollan los pasos abordados en cada etapa y la evolución en la cantidad de variables consideradas:

1. Preprocesamiento (ETL: extracción, transformación y carga).
 - Lectura de información .gdb y transformación a formato .csv
 - Correlación entre tabla de datos y diccionario
2. Integridad aplicada: reducción **desde 396 variables hasta 65**
 - Selección de variables medidas entre 60 y 100 cm (profundidad a la cual suelen disponerse los ductos enterrados)
 - Agrupación de variables por mecanismo de daño: corrosión, fuerzas ambientales, daño por terceros (**Tabla 1**)

Tabla 1. Cantidad de variables seleccionadas con criterio ingenieril y clasificación según su función o mecanismo de daño.

Mecanismo de daño o función	Cantidad de variables
Corrosión	46
Fallas por Corrosión	1
Susceptibilidad a fuerzas ambientales	8
Fallas por fuerzas ambientales	1
Daño por terceros	5
Fallas por daño por terceros	2
Consecuencia: valor económico	2
Total general	65

3. Limitación modelo a corrosión: reducción **desde 65 variables hasta 47**
 - Se decidió limitar el modelado a corrosión dado que es el principal mecanismo de daño del dataset.
4. Estadística aplicada: reducción **desde 47 variables hasta 26**
 - **Extracción** de variables con más del 20% de valores **nulos**, reduciendo **hasta 45 variables** asociadas a corrosión.

- Imputación de mediana a vacíos en variables numéricas e imputación de “desconocido” a vacíos en variables categóricas.
 - Clasificación de variables de fallas: se generaron variables binarias adoptando el valor 1 si se registraron fallas y 0 en caso contrario.
 - Análisis de correlación y **descarte de variables con correlación > 0.85** mediante biblioteca *seaborn* (**Figura 1**), reduciendo **hasta 42 variables** asociadas a corrosión.
 - Depuración de variables con índice $VIX > 10$ (Variance Inflation Factor) para evitar redundancia (multicolinealidad) mediante biblioteca *statsmodels* (**Figura 2**) reduciendo **hasta 25 variables** asociadas a corrosión.
5. Partición del data set para el entrenamiento y la validación de modelos
 - Partición del data set en muestra de entrenamiento (80%) y muestra de prueba (20%).
 6. Implementación de modelos de estadística clásica
 - Modelado Poisson mediante biblioteca *statsmodels*.
 - Modelado logístico mediante biblioteca *statsmodels*.
 7. Implementación de modelos de aprendizaje automático
 - Modelado Random Forest mediante biblioteca *sklearn*.
 8. Cálculo de riesgo y priorización de segmentos
 - Para cada segmento bajo análisis se calculó la consecuencia, el riesgo y se ordenaron de manera decreciente.

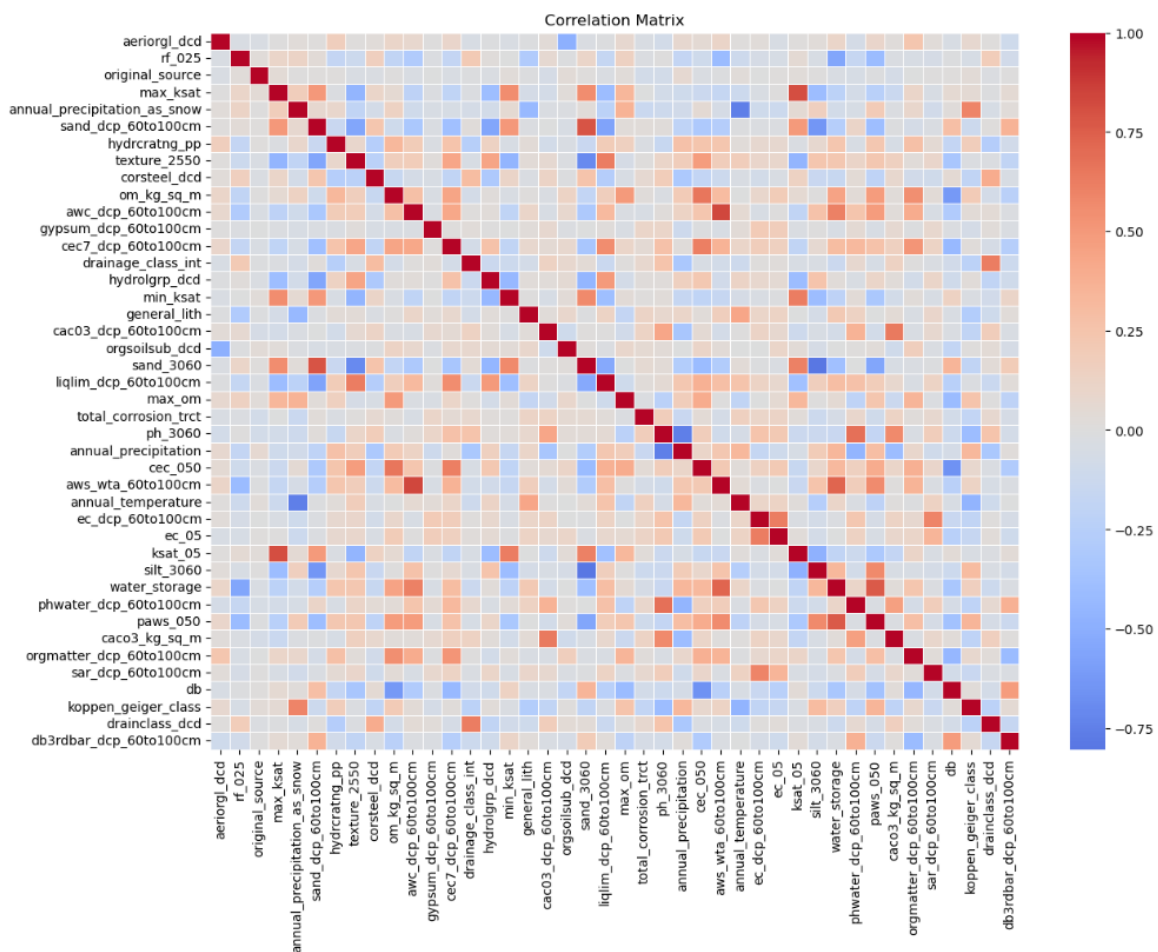


Figura 1. Matriz de correlación entre variables asignadas al mecanismo de corrosión, generado con biblioteca *seaborn*.

```
Eliminando 'db' con VIF = 282.15
Eliminando 'ph_3060' con VIF = 157.34
Eliminando 'db3rdbar_dcp_60to100cm' con VIF = 81.16
Eliminando 'paws_050' con VIF = 73.08
Eliminando 'aeriorsl_dcd' con VIF = 60.11
Eliminando 'phwater_dcp_60to100cm' con VIF = 57.88
Eliminando 'annual_temperature' con VIF = 36.80
Eliminando 'awc_dcp_60to100cm' con VIF = 36.28
Eliminando 'general_lith' con VIF = 29.99
Eliminando 'water_storage' con VIF = 28.18
Eliminando 'sand_3060' con VIF = 21.01
Eliminando 'liqlim_dcp_60to100cm' con VIF = 19.15
Eliminando 'koppen_geiger_class' con VIF = 17.89
Eliminando 'mean_ksat' con VIF = 15.27
Eliminando 'texture_2550' con VIF = 11.92
Eliminando 'drainage_class_int' con VIF = 11.85
Eliminando 'aws_wta_60to100cm' con VIF = 10.75
```

Figura 2: Variables descartadas como resultado de proceso iterativo usando biblioteca *statsmodels* para evitar redundancia mediante depuración de variables usando índice VIF.

2.3 Metodología estadística clásica

El análisis de los datos de fallas en ductos presenta particularidades que condicionan la elección del modelo estadístico dado que requiere modelos capaces de capturar la naturaleza discreta y de los datos disponibles.

En este data set, las variables de interés (variables explicadas) corresponden a conteos de fallas acumuladas en áreas geográficas específicas, en tanto que las variables explicativas corresponden a características ambientales y del entorno. Es decir, **cada observación representa un punto de agregación donde se registran las propiedades del suelo, condiciones climáticas y, además, el número de fallas ocurridas en la vecindad. Estos conteos no son valores promediados ni globales, sino acumulaciones locales que reflejan la intensidad de eventos de integridad y pueden dar cuenta de la susceptibilidad a la falla.** Por eso, para seleccionar la metodología estadística apropiada para este conjunto de datos se presentan las siguientes consideraciones:

- Que las variables de interés sean conteos de fallas implica que no pueden modelarse adecuadamente con métodos lineales clásicos (como regresión lineal). Estos modelos asumen una variable dependiente continua, normalmente distribuida lo cual no tiene sentido en el contexto de conteos. Por ejemplo, no es posible tener “-2 fallas” o “0,5 fallas”. Además, los conteos suelen presentar distribuciones asimétricas, lo que rompe los supuestos de normalidad y homocedasticidad de los modelos lineales.
- El data set exhibe una proporción elevada de ceros ($\approx 60\%$), que podría modelarse con un modelo de Poisson si las variables explicativas son estadísticamente significativas.
- Es posible transformar la variable de interés a una variable binaria que vale 1 si hay conteo de falla y 0 si no hay conteo de fallas previas. Luego, el fenómeno de falla podría modelarse mediante una regresión logística si las variables explicativas son estadísticamente significativas.

2.4 Metodología machine learning: Random Forest

El Random Forest [17] es un método de aprendizaje supervisado basado en el ensamble de múltiples árboles de decisión. Su principio operativo consiste en construir cientos de árboles de manera independiente, donde cada uno se entrena con una submuestra aleatoria de los datos (técnica conocida como bagging, **Figura 3**). Luego, se promedian las predicciones de todos los árboles individuales, lo cual permite reducir la varianza. Este método se caracteriza por su capacidad para capturar no-linealidades, umbrales críticos, interacciones complejas entre las variables.

Este método puede implementarse con tantas ramificaciones como variables existan, o bien limitando la “profundidad” del árbol. Es decir, se puede limitar la complejidad del modelo para evitar sobreajuste (y la consecuente pérdida de generalización) y reducir el esfuerzo computacional.

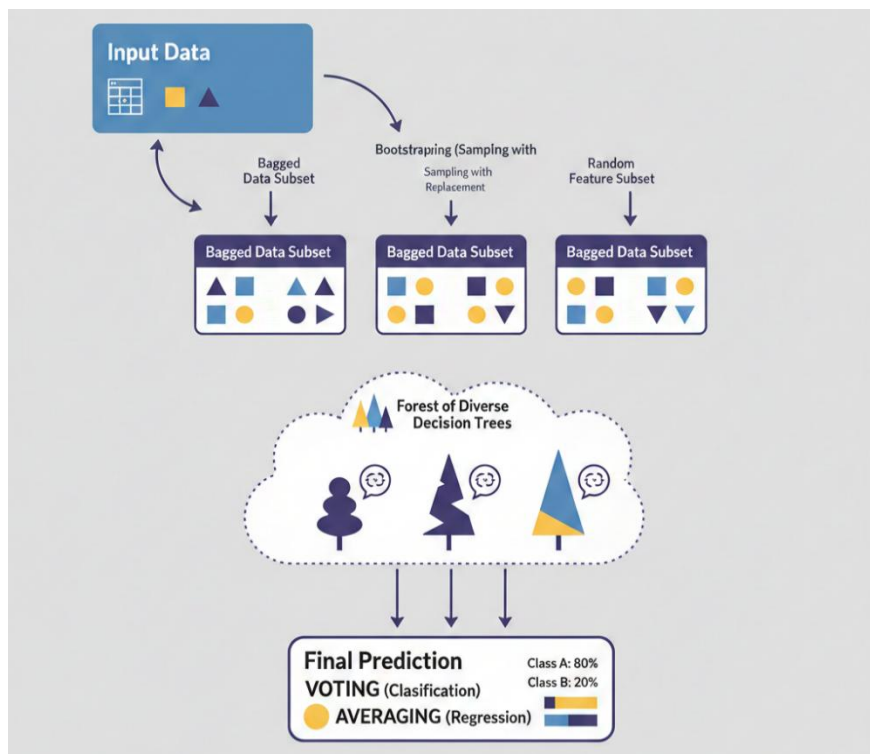


Figura 3: Esquema conceptual de modelo Random Forest.

2.5 Métricas para comparación de modelos

A continuación, se describen las métricas utilizadas para evaluar los modelos predictivos de clasificación (Falla vs. No Falla):

- Exactitud (Accuracy): porcentaje total de predicciones correctas.
- Precisión: grado de proximidad, repetibilidad o concordancia entre múltiples predicciones obtenidas.
- Sensibilidad (Recall): Responde a: De todas las fallas que ocurrieron, ¿cuántas fue capaz de detectar el modelo? En integridad, un Recall alto es vital para no omitir amenazas reales (falsos negativos).

- Puntuación F1 (F1 score): media armónica entre la Precisión y la sensibilidad. Es un indicador crítico cuando se busca un equilibrio en la predicción y las clases están desbalanceadas (como el data set bajo análisis, con alto contenido de valores nulos)
- Matriz de Confusión: es una herramienta visual que permite desglosar el desempeño del modelo entre Verdaderos Positivos, Verdaderos Negativos, Falsos Positivos y Falsos Negativos (el escenario más peligroso; fallas no detectadas que pueden comprometer la integridad).
- Curva ROC (Receiver Operating Characteristic) representa la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos a diferentes umbrales.
- AUC (Área Bajo la Curva): Es el valor numérico que resume la curva ROC. Un AUC de 0.5 equivale a una predicción al azar, mientras que 1.0 es una predicción perfecta, es decir que un AUC alto indica que el modelo es altamente eficaz para jerarquizar segmentos de mayor riesgo frente a los de menor riesgo, permitiendo una asignación de recursos eficiente

2.6 Resultados y comparación de modelos

A continuación, se comparan los resultados de los modelos implementados utilizando las métricas descriptas.

Tabla 2. Métricas obtenidas con modelos de estadística clásica y de machine learning aplicados al data set de AIIM [1].

Modelo	Exactitud	Sensibilidad	Precisión	F1-Score	AUC-ROC
Poisson	0.530	0.889	0.255	0.397	0.753
Random Forest (conteo)	0.675	0.868	0.334	0.482	0.850
Regresión Logística (Binario)	0.706	0.655	0.327	0.436	0.764
Random Forest (binario)	0.899	0.936	0.645	0.763	0.972
Random Forest limitado (binario)	0.786	0.819	0.438	0.571	0.880

Los resultados muestran que:

- El modelo de Poisson presenta los menores valores para casi todas las métricas analizadas. Sin embargo, presenta el segundo mejor valor de sensibilidad.
- La regresión logística ofrece una mejora en todas las métricas respecto a Poisson, pero a expensas de una pérdida de sensibilidad.
- Todos los modelos Random Forest presentaron mejoras en las métricas respecto a los modelos estadísticos clásicos.
- El modelo Random Forest binario presenta los mayores valores para todas las métricas analizadas.
- Limitar el tamaño de los árboles del modelo Random Forest binario brinda muy buenas predicciones, presentando el caso “second best” para 4 de las 5 métricas analizadas.

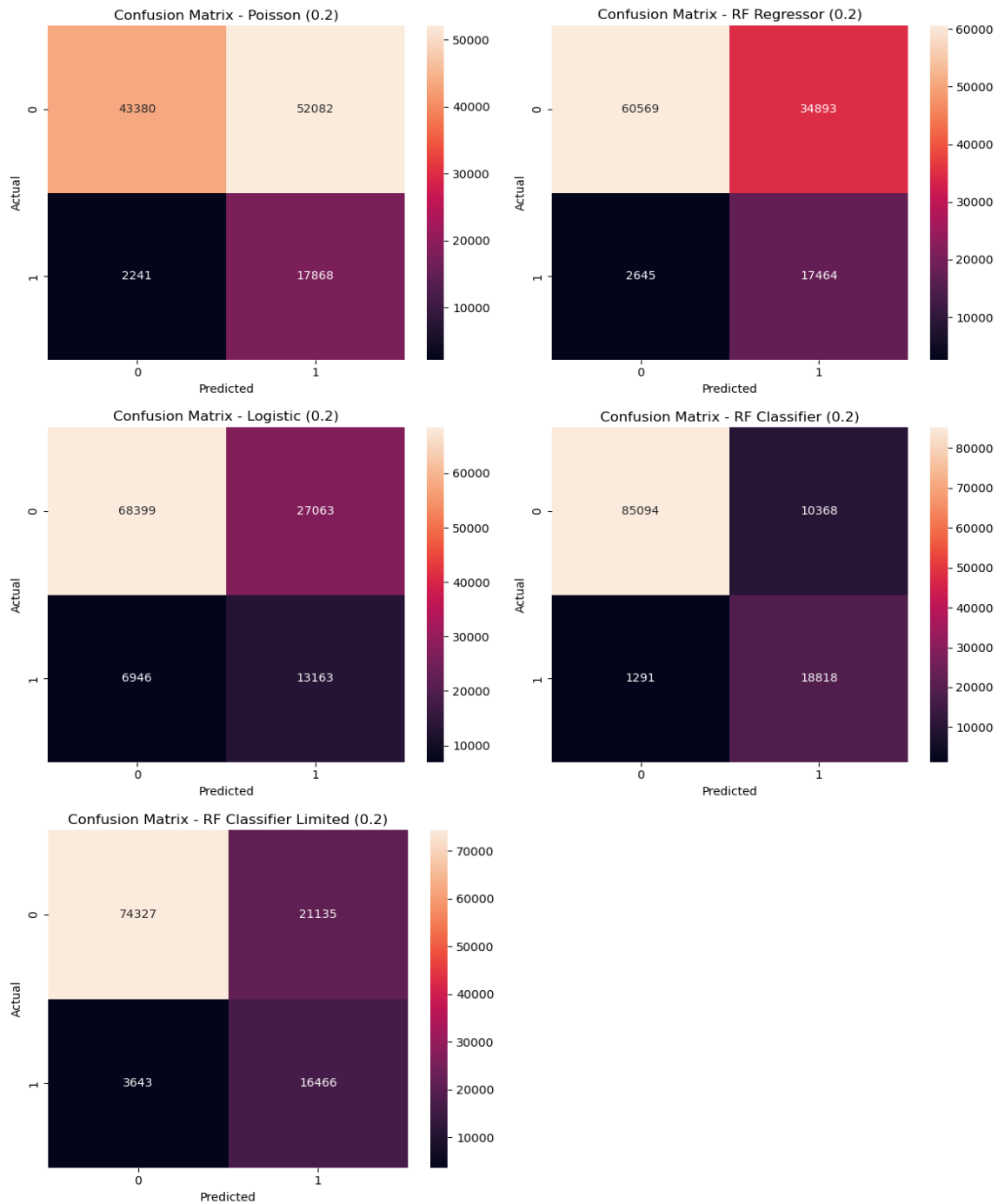


Figura 4. Matrices de confusión obtenidas con modelos de estadística clásica y de machine learning aplicados al data set de AIIM [1].

Las matrices de confusión nos permiten profundizar el análisis de los resultados. En el caso bajo análisis se observa que:

- El modelo Poisson predice mas falsos negativos que negativos reales (lo cual explica su baja precisión). Por otro lado, es adecuado prediciendo fallas (positivos reales).
- Obtener predicciones de conteo mediante Random Forest revierte el problema de falsos negativos del modelo Poisson, y mantiene el adecuado desempeño en la predicción de fallas.
- Todos los modelos de clasificación (falla/no falla) presentan mejor desempeño (precisión) en la identificación de negativos reales.

- La menor cantidad de falsos negativos se obtuvo con el modelo de clasificación Random Forest. A su vez, este modelo obtuvo la mayor precisión en la identificación de fallas. Esta combinación explica por qué presentan el F1-score más alto.

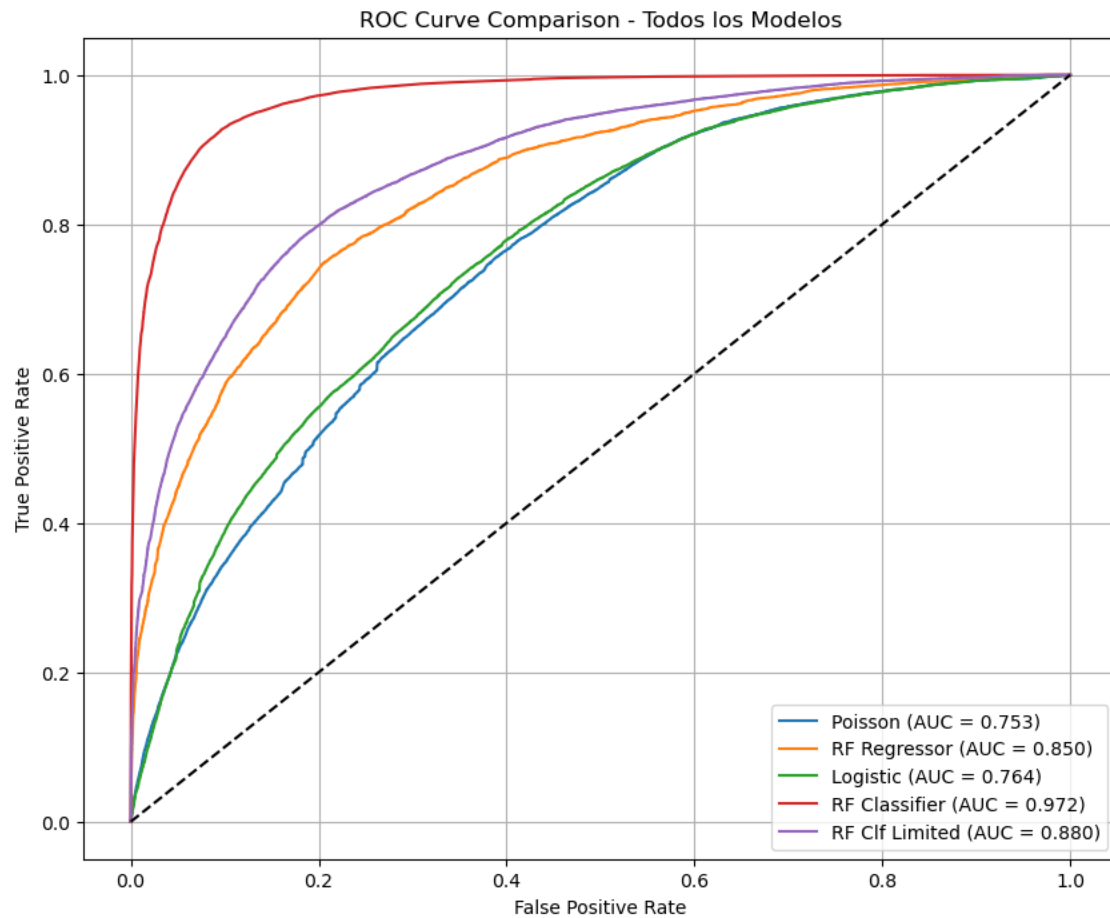


Figura 5. Curvas ROC obtenidas con modelos de estadística clásica y de machine learning aplicados al data set de AIIM [1].

Como se puede observar en las curvas ROC, las predicciones clásicas ofrecen buenos resultados (alejados de la diagonal que representa un resultado aleatorio). El modelo de regresión logística presenta una leve mejora en la predicción respecto al modelo de Poisson. Por otro lado, los modelos Random Forest permiten obtener mejores predicciones que los modelos de estadística clásica a lo largo de toda la curva, obteniendo los mejores resultados con el modelo de clasificación sin limitación de profundidad.

2.7 Caso de negocio: beneficios cuantitativos

En esta instancia del desarrollo, hemos creado y evaluado distintos modelos predictivos que nos permiten cuantificar la probabilidad de ocurrencia de un evento de falla dadas ciertas condiciones ambientales, definidas por las variables que se encuentran en la base de datos AIIM. Para cuantificar el impacto de esas fallas, recurrimos al concepto de riesgo, definido como:

$$\text{Riesgo} = \text{probabilidad} \times \text{consecuencia}$$

En cuanto a las consecuencias, podemos aprovechar los datos “buidvalue” y “agrivalue” incluidos en AIIM y definidos como:

- “Buildvalue”: El valor de la propiedad (USD) del U.S. Federal Emergency Management Agency (FEMA) National Risk Index (NRI) dataset, por unidad de agregación (census tract).
- “Agrivalue”: El valor agrícola (USD) del U.S. Federal Emergency Management Agency (FEMA) National Risk Index (NRI) dataset, por unidad de agregación (census tract).

Es decir, que podemos obtener un indicador de consecuencias sumando “buidvalue” y “agrivalue”.

Luego, es posible determinar un valor de riesgo asociado a corrosión para cada uno de los modelos desarrollados y lo podemos comparar con el riesgo real, que surge de utilizar el recuento de fallas por corrosión (o bien, la variable binaria asociada: falla/no falla).

Con esta información, el caso de negocio consiste en ordenar el riesgo de cada unidad de agregación (región que contiene un ducto) de mayor a menor para cada uno de los modelos, y establecer dicho orden como criterio para la asignación de recursos. Así, se cuantifica qué fracción del riesgo real sería capturado si utilizamos los modelos.

A continuación, se presentan los resultados:

Tabla 3. Caso de negocio: fracción del riesgo económico real capturado si se asignan recursos a las unidades de agregación (segmentos de ductos) respetando orden de riesgo por corrosión, aplicado al data set de AIIM [1].

	Fracción del riesgo económico real capturado				
X% de activos más riesgosos	10%	20%	30%	40%	50%
Modelo					
Poisson	0.502	0.653	0.759	0.847	0.912
Random Forest (conteo)	0.691	0.812	0.883	0.929	0.957
Regresión Logística (Binario)	0.473	0.642	0.770	0.852	0.913
Random Forest (binario)	0.750	0.926	0.980	0.992	0.997
Random Forest limitado (binario)	0.634	0.789	0.877	0.930	0.960
Resultado real	0.886	1.000	1.000	1.000	1.000

Los resultados del caso de negocio confirman las conclusiones obtenidas mediante las métricas estudiadas previamente: el modelo Random Forest binario (“RF_clasifier”) presenta las mejores predicciones.

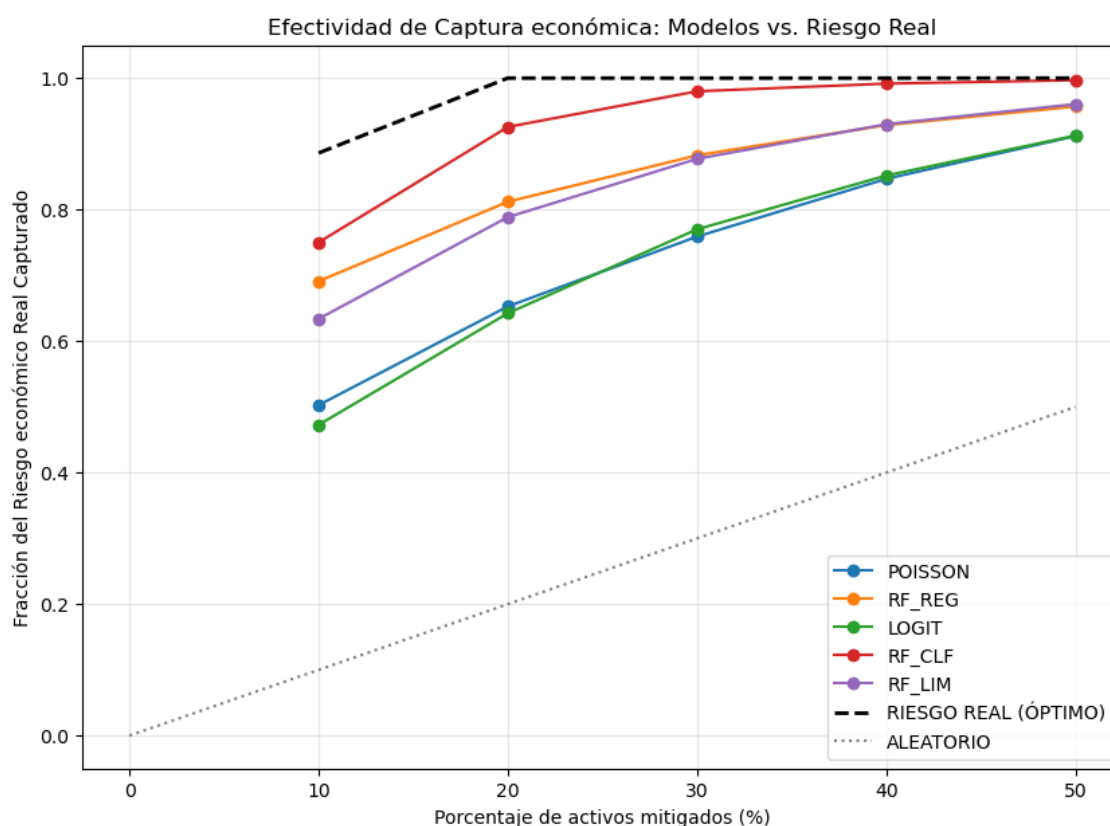


Figura 6. Curva de ganancia económica: efectividad en captura de riesgo económico en función del modelo de riesgo utilizado, aplicado al data set de AIIM [1].

Conclusiones

El presente trabajo demuestra que **la transición de modelos de integridad tradicionales hacia modelos proactivos basados en Ciencia de Datos no solo es técnicamente factible, sino económicamente superior.** A partir de los resultados obtenidos, se concluye lo siguiente:

- Es posible generar conocimiento y optimizar la toma de decisiones de integridad y gestión de activos creando modelos predictivos basados en aprendizaje automático a partir de bases de datos históricas, ya sean públicas o privadas.
- Los modelos desarrollados permitieron procesar 577.852 registros y 396 variables operativas y ambientales de la base AIIM, transformando datos históricos en conocimiento predictivo.
- Se validó la existencia de una alta concentración de riesgo (fenómeno de Pareto), donde la totalidad del impacto económico observado en la base de datos AIIM se localiza en una fracción minoritaria de los activos. **Esto sugiere que los planes de inspección basados en plazos predeterminados o en frecuencias de falla provenientes de bases de datos pueden incurrir en una sobre-inspección de activos sanos y una sub-estimación de puntos críticos.**
- La comparación de modelos reveló que **el Random Forest Classifier presenta el mejor desempeño para la toma de decisiones operativas.** Este modelo superó a los métodos econométricos tradicionales (Logit y Poisson) al capturar un 92.5% del riesgo económico real inspeccionando únicamente el 20% de los activos.
- Se determinó que es posible obtener buenas predicciones con modelos binarios (regresión logística) y también con modelos de conteo (Poisson/Regresión). En todos los casos es

posible optimizar las predicciones con tecnologías de aprendizaje automático (machine learning), específicamente en este caso modelos Random Forest.

- son óptimos para el ranking de prioridad de intervención, Desde una perspectiva económica, la implementación de este sistema permite optimizar la asignación de recursos de mantenimiento hacia las áreas donde el Riesgo Capturado (USD) es mayor. Esto reduce drásticamente el costo de oportunidad asociado a fallas no detectadas y fallos en la asignación de presupuestos.

Bibliografía

[1] Lucy Romeo, Devin Justman, Alec Dyer, Caleb Malay, Catherine Schooley, Jennifer Bauer, Dakota Zaengle, Kelly Rose, Advanced Infrastructure Integrity Modeling (AIIM) Onshore Pipeline Database, 6/9/2024, <https://edx.netl.doe.gov/dataset/advanced-infrastructure-integrity-modeling-aiim-onshore-pipeline-database>, DOI: 10.18141/2371242.

[2] NACE International, NACE SP0502, “Pipeline External Corrosion Direct Assessment Methodology”, https://doi.org/10.5006/NACE_SP0502-2025.

[3] NACE International, “NACE SP0206 Internal Corrosion Direct Assessment for pipelines carrying normally Dry Natural Gas”, https://doi.org/10.5006/NACE_SP0206-2016.

[4] The American Society of Mechanical Engineers, “ASME B31.8S Managing System Integrity of Gas Pipelines”.

[5] The American Petroleum Institute, “API RP 1160 Managing System Integrity for Hazardous Liquid Pipelines”.

[6] Muhlbauer, W. Kent, “Pipeline risk management manual: a tested and proven system to prevent loss and assess risk”.

[7] British Standards Institution, “PD 8010-3:2009+A1:2013 Pipeline systems. Steel pipelines on land. Guide to the application of pipeline risk assessment to proposed developments in the vicinity of major accident hazard pipelines containing flammables. Supplement to PD 8010-1:2004”.

[8] Institution of Gas Engineers and Managers, IGEM/TD/2 “*Assessing the risks from high pressure Natural Gas pipelines* (IGEM/TD/2 Edition 2).

[9] Pipeline and Hazardous Materials Safety Administration. “*Pipeline Incident Flagged Files*. U.S. Department of Transportation”, <https://www.phmsa.dot.gov/data-and-statistics/pipeline/pipeline-incident-flagged-files>.

[10] United Kingdom Onshore Pipeline Association, “*UKOPA Pipeline Fault Database: Product Loss Incidents and Faults Report (1962 – 2022)* (Report No. UKOPA/RP/24/001)”.

[11] Instituto Argentino del Petróleo y del Gas, “*RIDA - Registro de Incidentes de Ductos de la Argentina: Informe estadístico anual*. Comisión de Transporte de Hidrocarburos”.

[12] The American Petroleum Institute, API Recommender Practice 580, “Elements of a Risk-Based Inspection Program”

[13] The American Petroleum Institute, API Recommender Practice 581, “Risk-Based Inspection Methodology”

[14] The International Organization for Standardization, ISO 55000, “Asset management – Overview, principles and terminology”.

[15] The International Organization for Standardization, ISO 55001, “Asset management — Asset management system — Requirements”.

[16] The International Organization for Standardization, ISO 55002, “Asset management — Management systems — Guidelines for the application of ISO 55001”.

[17] L. Breiman, “Random Forests”, Machine Learning 45, 5-32 (2001) <https://doi.org/10.1023/A:1010933404324>.

CV Diego Sebastián López

Título y lugar de estudio: Ingeniero en Materiales – Universidad Nacional de Mar del Plata, candidato a master en economía aplicada – Universidad Torcuato Di Tella.

Trayectoria: 8 años de experiencia en gestión de integridad de activos en el sector energético e industrial trabajando en Camuzzi, en PAE y como consultor profesional. Especializado en el diseño de estrategias de operación, mantenimiento y gestión de activos para optimizar la asignación de recursos y agilizar la toma de decisiones en función de los riesgos. Además, aporta una visión estratégica, integrando modelos técnico-económicos y gestión de datos en las decisiones de ingeniería que permiten maximizar la confiabilidad, optimizando asignación de recursos y garantizando el cumplimiento normativo.

Cargo actual en la empresa: Consultor independiente en Ingeniería, integridad y optimización de mantenimiento, ofreciendo soluciones integrales para la gestión de activos y la toma de decisiones operativas.