



**UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA CIVIL**



**Identificación de variables influyentes en calificación de condición de deterioro de puentes
usando algoritmos de *Machine Learning***

POR

Lorenzo Reyes Pedernera

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción para
optar al título de Ingeniero Civil

Profesor Guía
Tomás Echaveguren Navarro

Abril 2024
Concepción (Chile)

© 2024 Lorenzo Reyes Pedernera

© 2024 Lorenzo Reyes Pedernera

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

DEDICATORIA

*Para mi familia, que me enseñó el esfuerzo
y la dedicación como herramientas para generar cambios.*

RESUMEN

Chile cuenta con bases de datos detalladas de tráfico de vehículos y catastro e inspección de puentes que no se encuentran conectadas entre sí. En la inspección de puentes se incluye la calificación de condición de deterioro de puentes, usada para tomar decisiones respecto a la mantención de puentes, pero no predice la condición de puentes en el tiempo. Para la calibración de un modelo de comportamiento de puentes es necesario determinar qué variables incluir en este, ya que la inclusión o exclusión de variables explicativas significativas o ponderar erróneamente estas variables podría afectar en la habilidad de predicción del modelo.

El propósito de la memoria fue determinar la influencia de variables explicativas en la calificación de condición de puentes usando bases de datos en Chile. Se midió la influencia de las variables utilizando ReliefF e Importancia de Atributos por Permutación aplicada en las técnicas *Random Forest* y *Extreme Gradient Boosting*.

Se concluyó que las variables más influyentes fueron la cantidad de deterioros en un puente, la longitud total del puente, el ancho total del puente, la cantidad de deterioros por la longitud del puente y la materialidad de las vigas. De las variables influyentes, solo la cantidad de deterioros y la causa de deterioros se encuentran incluidas en la calificación de condición actual. Estas variables pueden orientar la calibración de modelos de comportamiento de puentes.

ABSTRACT

Chile has detailed databases of traffic and bridge inventory and inspection that are not connected. The bridge inspection includes the bridge condition rating, used to make maintenance decisions but it cannot predict the bridge condition in time. To calibrate a bridge condition behavior model, it is crucial to determine which explanatory variables need to be considered, as including or excluding significant explanatory variables or weighting variables wrongfully may affect the prediction ability of the behavior model.

The purpose of this research was to determine the influence of explanatory variables in the bridge condition rating using Chilean databases. The influence of variables was measured using ReliefF and Feature Importance by Permutation joined with Random Forest and Extreme Gradient Boosting.

In conclusion the most influential variables weighted by percentage score were the total number of deteriorations in the bridge, total bridge length, total bridge width, number of deteriorations by bridge longitude and girder material. Among the influential variables only the total number of deteriorations and the cause of deterioration are included in the bridge condition rating calculations. These variables can inform the calibration of bridge condition behavior models.

AGRADECIMIENTOS

Agradezco a mi familia por su apoyo incondicional para enfocarme en mis estudios, acompañándome en mi aprendizaje, a mis profesores por dedicarle tiempo a cada duda y cuestionamiento que tuve al aprender, a mi profesor guía Tomás Echaveguren por su apoyo en ampliar las técnicas aplicadas a ingeniería civil y a la Universidad por darme las herramientas necesarias en la búsqueda de generar soluciones desde el conocimiento.

ÍNDICE DE CONTENIDOS

1	INTRODUCCIÓN	1
1.1	Motivación	1
1.2	Objetivo general	2
1.3	Objetivos específicos.....	2
1.4	Metodología de trabajo.....	2
1.5	Principales resultados y conclusiones	3
1.6	Organización de documento.....	3
2	INVENTARIO E INSPECCIÓN DE PUENTES	5
2.1	Introducción	5
2.2	Inventario de puentes en Chile	5
2.3	Inspección de puentes en Chile	5
2.4	Calificación de condición de puentes en Chile	6
2.5	Conclusiones	7
3	HERRAMIENTAS DE MACHINE LEARNING PARA LA MEDICIÓN DE INFLUENCIA DE VARIABLES EXPLICATIVAS.....	9
3.1	Introducción	9
3.2	Clasificación de técnicas de <i>Machine Learning</i>	9
3.3	Aplicaciones de ML en medición de influencia de atributos	13
3.4	Conclusiones	14
4	METODOLOGÍA DE ANÁLISIS CON HERRAMIENTAS DE MACHINE LEARNING....	16
4.1	Introducción	16
4.2	Diagrama de relación de actividades por objetivo específico	16
4.3	Configuración de base de datos.....	17
4.4	Selección de técnicas de ML.....	17

4.5	Aplicación de técnicas de ML.....	18
4.6	Conclusiones	19
5	CONFIGURACIÓN DE BASE DE DATOS.....	20
5.1	Introducción	20
5.2	Unificación de base de datos.....	20
5.3	Estadística descriptiva de variables.....	21
5.4	Variables incluidas en el análisis	34
5.5	Depuración de bases de datos unificadas	37
5.6	Conclusiones	38
6	ANÁLISIS CON HERRAMIENTAS DE MACHINE LEARNING	40
6.1	Introducción	40
6.2	Criterios de selección de técnicas.....	40
6.3	Técnicas seleccionadas.....	40
6.4	Aplicación de técnicas.....	42
6.5	Influencia relativa de variables	44
6.6	Conclusiones	57
7	CONCLUSIONES Y RECOMENDACIONES.....	59
	REFERENCIAS	61
	ANEXOS.....	63
	ANEXO 2.1 Contenido de variables de tráfico vial, inventario e inspección de puentes.....	65
	ANEXO 2.2 Cálculo de calificación de condición de puentes	74
	ANEXO 5.1 Estadística descriptiva de variables.....	78
	ANEXO 6.1 Ranking de influencia de variables	83

ÍNDICE DE TABLAS

Tabla 2.1 Categorías de calificación de condición de puentes.....	6
Tabla 3.1 Catálogo de técnicas de Machine Learning	10
Tabla 5.1 Variables en bases de datos de catastro e inspección de puentes presentes	21
Tabla 5.2 Estadística descriptiva por Tipología Estructural – Región zona centro	22
Tabla 5.3 Estadística descriptiva por Tipología Estructural – Región zona sur	24
Tabla 5.4 Ranking de Tipos de deterioro recurrentes – Región zona centro	30
Tabla 5.5 Distribución de Gravedad y Extensión en deterioros.....	31
Tabla 5.6 Descripción de variables creadas	35
Tabla 5.7 Variables incluidas en la modelación.....	36
Tabla 5.8 Cantidad de registros excluidos en depuración Región Zona Sur.....	38
Tabla 6.1 Rendimiento de técnicas en regresión y clasificación por región	44
Tabla A.2.1.1 Contenido de variables de Tráfico	65
Tabla A.2.1.2 Contenido de variables de inventario de puentes	65
Tabla A.2.1.3 Contenido de variables de inspección de puentes	69
Tabla A2.2.1 Ponderación de Extensión y Gravedad en Calificación de Deterioro I3MOP	75
Tabla A2.2.2 Ponderación de Tipo de Deterioro en Calificación de Deterioro I3MOP	75
Tabla A2.2.3 Ponderación de Causa de Deterioro en Calificación de Deterioro I3MOP	76
Tabla A2.2.4 Ponderación de Unidad en Calificación de Deterioro I3MOP	77
Tabla A5.1 Registros vacíos de variables originales – Región Zona Centro	78
Tabla A5.2 Registros vacíos de variables originales – Región Zona Sur	80
Tabla A5.3 Categorías de deterioro.....	81

ÍNDICE DE FIGURAS

Figura 2.1 Esquema de procedimiento de cálculo de calificación de condición de puentes.....	7
Figura 3.1 Clasificación de técnicas de Machine Learning	10
Figura 4.1 Diagrama de relación de actividades	16
Figura 5.1: Distribución de calificación de condición por Tipología Estructural – Región zona centro	23
Figura 5.2: Distribución de calificación de condición por Tipología Estructural – Región zona sur	25
Figura 5.3: Distribución de calificación de condición por material de cepas – Región zona centro .	26
Figura 5.4: Distribución de calificación de condición por material de cepas – Región zona sur	26
Figura 5.5: Distribución de calificación de condición por material de vigas – Región zona centro.	27
Figura 5.6: Distribución de calificación de condición por material de vigas – Región zona sur.....	28
Figura 5.7: Distribución de calificación de condición por material de losas – Región zona centro ..	28
Figura 5.8: Distribución de calificación de condición por material de losas – Región zona sur	29
Figura 5.9 Calificación de condición por Gravedad – Región zona centro	32
Figura 5.10 Calificación de condición por Gravedad – Región zona sur	32
Figura 5.11 Calificación de condición por Extensión – Región zona centro.....	33
Figura 5.12 Calificación de condición por Extensión – Región zona sur.....	33
Figura 6.1 Implementación de técnicas en Orange	43
Figura 6.2 Ranking de variables influyentes obtenidas mediante ReliefF.....	45
Figura 6.3 Ranking de variables influyentes obtenidas mediante Random Forest (en tareas de clasificación)	47
Figura 6.4 Ranking de variables influyentes obtenidas mediante XGBoost (en tareas de clasificación)	48
Figura 6.5 Ranking de variables influyentes obtenidas mediante Random Forest (en tareas de regresión).....	50
Figura 6.6 Ranking de variables influyentes obtenidas mediante XGBoost (en tareas de regresión)	51
Figura 6.7 Ranking de variables filtradas (clasificación) - Región zona centro	53
Figura 6.8 Ranking de variables filtradas (clasificación) - Región zona sur	54
Figura 6.9 Ranking de variables filtradas (regresión) - Región zona centro.....	55
Figura 6.10 Ranking de variables filtradas (regresión) - Región zona sur.....	55
Figura 6.11 Influencia promedio de variables principales	56

Figura A6.1.1 Ranking de Variables Filtradas (Clasificación) - Región Zona Centro	83
Figura A6.1.2 Ranking de Variables Filtradas (Clasificación) - Región Zona Sur	84
Figura A6.1.3 Ranking de Variables Filtradas (Regresión) - Región Zona Centro	85
Figura A6.1.4 Ranking de Variables Filtradas (Regresión) - Región Zona Sur	85

1 INTRODUCCIÓN

1.1 Motivación

Los puentes son activos críticos en el servicio de infraestructura vial de Chile debido a que su fallo puede dejar sin conexión a ciudades completas. La Dirección de Vialidad de Chile tiene bajo su tuición 6781 puentes y utiliza la aplicación INSPECTOR para auscultar sus deterioros y generar una calificación de condición global, ponderando características de los deterioros presentes como área que abarcan, severidad del daño, causa del deterioro y elemento afectado en su cálculo.

En Chile no existe un modelo de comportamiento de puentes. La calificación de condición solo muestra el estado del puente en un momento específico y no proyecta su condición en el tiempo. La Dirección de Vialidad cuenta con bases de datos de tráfico de vehículos en la red vial y de inventario e inspección de puentes. Las bases de datos detallan el TMDA por tipo de vehículo en la red vial, las características físicas de los puentes y distintas características de sus deterioros como su tipo, extensión y severidad. Es necesario determinar la influencia de estas variables en la condición de los puentes para calibrar y validar un modelo de comportamiento ya que la inclusión o exclusión de variables significativas podría afectar funcionamiento de este.

La relación entre la calificación de condición y las distintas variables explicativas es compleja debido a la cantidad de variables numéricas, categóricas ordinales y categóricas nominales involucradas simultáneamente. Además, las bases de datos de la Dirección de Vialidad presentan miles de puentes auscultados. En este caso las técnicas de estadística convencionales pierden interpretabilidad. Las herramientas de *Machine Learning* (ML) se adaptan a problemas con distintitos tipos de variables y permiten clasificar o predecir fenómenos a partir del aprendizaje de patrones en bases de datos. Debido a las características de los datos disponibles y de la cantidad de puentes auscultados se propone el uso de ML para el análisis de influencia de variables explicativas.

Esta memoria de título tiene por objetivo determinar la influencia de distintas variables en la calificación de condición de puentes, usando bases de datos de puentes localizados en una región del centro y otra del sur de Chile. La influencia relativa de las variables puede orientar una actualización de modelos de condición de puentes, la calibración de modelos de predicción de comportamiento de

puentes o políticas de acción como el uso de estrategias de mantención futuras. El Anexo 1.1 muestra la contribución de la Memoria de Título a los Objetivos de Desarrollo Sostenible.

1.2 Objetivo general

Determinar la influencia de variables de dimensión, materialidad y tráfico en la calificación de condición de puentes en bases de datos en Chile.

1.3 Objetivos específicos

1. Integrar bases de datos de inventario e inspección de puentes y tráfico vial en regiones la zona centro y sur del país.
2. Seleccionar técnicas de *Machine Learning* para análisis de patrones acordes a las características de los datos disponibles.
3. Identificar la influencia relativa de variables en la calificación de condición, obtenida por las técnicas de *Machine Learning* seleccionadas.

1.4 Metodología de trabajo

Inicialmente se unificaron las bases de datos de tráfico y catastro e inspección de puentes de dos regiones del centro y del sur de Chile por región, analizando el contenido de sus variables presentes a través de estadística descriptiva. Luego se identificaron variables que tengan información similar que deba ser agregada, o bien la falta de conexión entre variables presentes, agregando variables nuevas que unifiquen esta semejanza o que relacionen las variables. Se excluyeron de la base de datos unificada las variables con más del 50% de registros sin información y los registros que no presenten la mínima cantidad de información necesaria para confeccionar las nuevas variables.

Se realizó una revisión exploratoria de publicaciones enfocadas en la determinación de influencia de variables en modelos explicativos, para luego comparar las ventajas y desventajas de las técnicas de ML disponibles. Fueron seleccionadas para el análisis las técnicas de ML que incluyan variables numéricas y categóricas simultáneamente en su modelación, que se adapten a la falta de relación directa entre variables dependientes e independientes, que presenten parámetros de modelación de baja complejidad y que se encuentren disponibles en software accesible.

Finalmente se generó un modelo de las técnicas de *Machine Learning* seleccionadas para la medición de influencia relativa de las variables en la calificación de condición de puentes. La medición de influencia de las variables se realizó en dos etapas, la primera incluye todas las variables explicativas. En la segunda etapa se descartan las variables menos influyentes en la primera etapa y se genera un ranking de variables según su influencia en la calificación de condición de puentes. Las mediciones de influencia difieren en escala para las distintas técnicas, por lo que se normalizó la influencia obtenida por técnica, para luego ser promediadas por variable, obteniendo la influencia general de la variable.

1.5 Principales resultados y conclusiones

Las cinco variables más influyentes fueron la cantidad de deterioros en un puente, la longitud total del puente, el ancho total del puente, la cantidad de deterioros por la longitud del puente y la materialidad de las vigas. La influencia de las variables decrece rápidamente entre los lugares del ranking, de la sexta a la décimo tercera posición la influencia de las variables es del 20% al 30% de la variable más influyente. Estas incluyen el número de cepas, el espesor de losa, la clase de ruta la cantidad de deterioros por superficie de tablero, la materialidad de la carpeta, la proporción de vehículos livianos del TMDA, el número de vigas y la materialidad de la losa.

De las variables influyentes, solo la cantidad de deterioros y la causa de deterioros se encuentran incluidas en el procedimiento de cálculo de la calificación de condición actualmente. Este ranking de variables influyentes en la calificación de condición de puentes puede orientar la calibración de modelos de comportamiento de puentes, tomando en cuenta que la cantidad de deterioros y las dimensiones del puente fueron más influyentes que las características de los deterioros, dando una ponderación mayor a estas variables en la predicción de comportamiento de los puentes.

1.6 Organización de documento

La memoria consta de siete capítulos. En el Capítulo 1 se introduce la motivación de identificar la influencia de distintas variables explicativas en la calificación de condición de puentes, el objetivo general y los objetivos específicos de la memoria, la metodología de trabajo para analizar las bases de datos y determinar la influencia de las variables con técnicas de ML y los principales resultados de influencia de variables. En el Capítulo 2 se contextualiza cómo son inventariados e inspeccionados

los puentes en Chile, en conjunto con el sistema de información que almacena estas bases de datos, para luego exponer el método de calificación de condición de estos. En el Capítulo 3 se describen distintas técnicas de *Machine Learning*, su funcionamiento y el tipo de datos de entrada y salida de estos. Se muestra además el trabajo de distintos autores que aplicaron técnicas de ML para la medición de influencia de variables en modelos explicativos en fenómenos de deterioro de activos.

El Capítulo 4 detalla el procedimiento de las actividades desarrolladas para alcanzar los objetivos de la memoria. Presenta una sección de cómo se configuran las bases de datos unificadas, los criterios de selección de variables y registros a incluir en el análisis, otra sección donde se detallan los criterios de selección de técnicas de ML y una última sección de cómo se aplican distintas técnicas de ML para la medición de influencia de variables en la memoria. En el Capítulo 5 se caracteriza el contenido de las bases de datos con estadística descriptiva, se crean nuevas variables y se selecciona las variables y registros a ser usados en la modelación de técnicas de ML. El Capítulo 6 presenta una sección de la selección de técnicas de ML a usar en el análisis, otra sección con la aplicación de las técnicas en software y una última sección con los resultados de medición de influencia de variables. Por último, en el Capítulo 7 se concluye sobre las principales variables influyentes en la calificación de condición de puentes, los objetivos de la memoria, sus limitaciones y las implicancias para futuros estudios.

2 INVENTARIO E INSPECCIÓN DE PUENTES

2.1 Introducción

En este capítulo se da un contexto de las variables incluidas en el inventario de puentes en Chile, cómo se inspeccionan y se detalla el procedimiento de calificación de condición de puentes utilizando actualmente por la Dirección de Vialidad del Ministerio de Obras Públicas [MOP].

2.2 Inventario de puentes en Chile

El inventario de puentes es generado a través de un catastro que recopila distintas características administrativas y técnicas. Algunas características administrativas son el nombre del puente, la codificación de identificación del puente, la ruta a la que pertenece, o su tuición. Algunas características técnicas son su tipología estructural y las dimensiones y materialidades de sus elementos (estribos, cepas, vigas, losas y carpeta). Las bases de datos con el catastro e inspección de puentes son almacenadas en el sistema de información I3MOP 2.0.

2.3 Inspección de puentes en Chile

La inspección de puentes en Chile es la auscultación de su estado de condición en un determinado momento, recopilando características de los deterioros presentes y generando una calificación de condición que es usada para tomar decisiones respecto al mantenimiento de puentes. La inspección es realizada a través de la aplicación INSPECTOR donde se ingresan datos de los deterioros presentes en un puente medidos en terreno.

La aplicación INSPECTOR requiere que se registren distintas variables de los deterioros en un puente para calcular su calificación de condición: el tipo y la causa del deterioro presente en cada elemento del puente (de un conjunto determinado de tipos y causas seleccionables), la gravedad del deterioro (en categorías leve, media, alta y muy alta) y medidas de la extensión del deterioro presente (difiere para cada tipo de deterioro, pueden ser variables de longitud, área, etc.). La aplicación asigna una codificación a cada deterioro registrado en el puente referenciando el elemento al que afecta. La información es almacenada en el Sistema de Información I3MOP 2.0.

El Plan de puentes 2020-2030 (Dirección de Vialidad MOP, 2022) plantea como uno de sus objetivos el optimizar el ciclo de vida de los puentes como activos de una red vial resiliente en etapas de

preinversión, inversión y operación, para lo cual propone generar un monitoreo permanente de los puentes. Es necesario orientar qué variables son influyentes para ser monitoreadas prioritariamente. De Solminihaç et al. (2018) clasifican los niveles de información para la gestión de activos de infraestructura vial según los niveles de calidad de información, describen que el nivel más detallado de información se usa en investigaciones detalladas de campo y en el diseño de proyectos en un corto plazo, mientras que en un segundo nivel de información menos detallado puede ser usado en horizontes temporales mayores orientando trabajos de mantenimiento.

2.4 Calificación de condición de puentes en Chile

La calificación de condición de puentes es una medida del deterioro del puente en un momento específico en el tiempo, representa el estado de los puentes en números enteros del 1 al 100, siendo el 1 el peor estado y el 100 el mejor estado (categorizadas en la Tabla 2.1). Es generada por la aplicación INSPECTOR a través de la ponderación de distintas variables de los deterioros auscultados en un puente como tipo de deterioro, causa de deterioro, extensión de deterioro y severidad de deterioro y elemento al que afecta el deterioro. Luego se selecciona la peor de las calificaciones de deterioro para una unidad y progresivamente se agregan y ponderan por la cantidad de unidades, elementos y componentes del puente. Esta es usada para tomar decisiones respecto a la mantención de los puentes.

Tabla 2.1 Categorías de calificación de condición de puentes (MOP, 2022)

Rango	Calificación de condición
80-100	Puente con deterioros de carácter durable o funcional leve
60-80	Puente con deterioros de carácter durable o funcional medios
40-60	Puente con deterioros de carácter resistente leves o de carácter durable o funcional extendidos
20-40	Puente con deterioros de carácter resistente medios. Deterioros durables o funcionales graves
1-20	Puente colapsado o con deterioros de carácter resistente muy graves

La Figura 2.1 muestra un esquema de cómo la aplicación INSPECTOR calcula la calificación de condición de puentes, se lee desde el lado izquierdo hacia el lado derecho, empezando por el ingreso de las características de tipo de deterioro, causa de deterioro, extensión del deterioro, gravedad del deterioro y elemento al que afecta el deterioro, para generar una calificación del deterioro en específico. Luego para cada uno de los deterioros en una Unidad se selecciona el deterioro con peor

calificación (calificación más baja del 1 al 100) y se aplican distintos métodos de agregación entre las Unidades, Elementos y Componentes hasta llegar a tener una calificación global del activo.

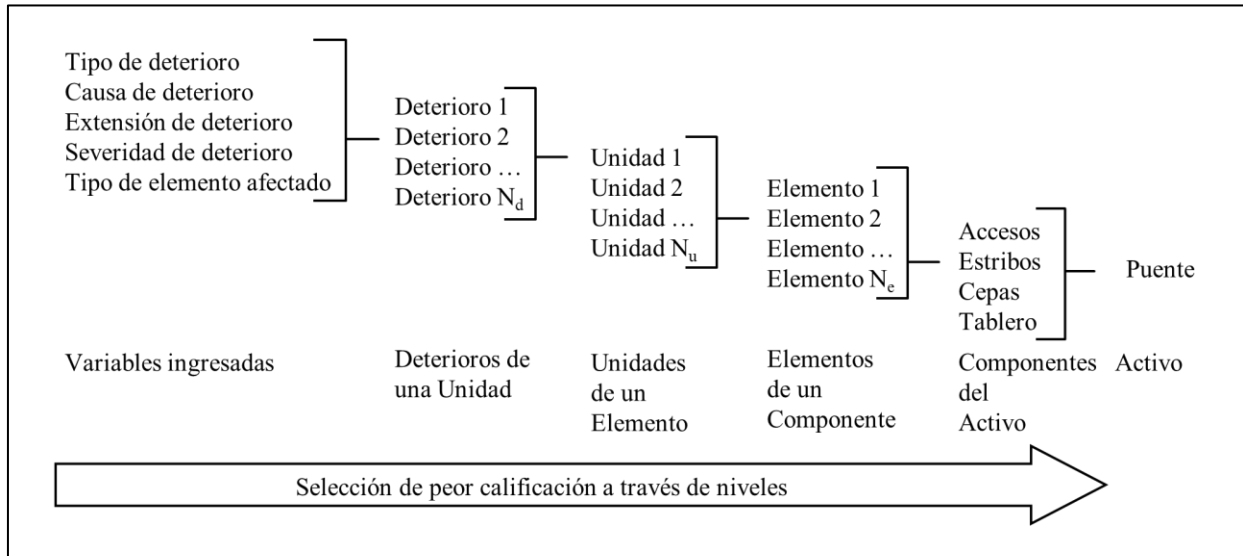


Figura 2.1 Esquema de procedimiento de cálculo de calificación de condición de puentes

2.5 Conclusiones

El sistema de información I3MOP almacena datos de las dimensiones de puentes, la materialidad de sus elementos, la antigüedad de los puentes, los deterioros presentes en un puente (incluyendo su tipo, extensión, severidad y causa) y la calificación de condición del puente con un nivel de información detallado. Sin embargo, no se incluyen otras variables importantes en modelos de comportamiento de puentes como las dimensiones de altura del puente en distintos puntos, las condiciones del terreno donde se encuentra emplazado, un historial de intervenciones de mantención o un registro de eventos adversos (ya sean naturales o antrópicos).

La inspección de puentes a través de la aplicación INSPECTOR permite almacenar los datos de inspección consistentemente, facilitando el análisis de la influencia de variables. El proceso de inspección no es completamente independiente del factor humano porque es posible que difieran resultados dependiendo del inspector por ejemplo en la estimación de la gravedad o la causa del deterioro.

El sistema de calificación de condición de puentes I3MOP se basa en las características de los deterioros (tipo, causa, extensión y gravedad) y la cantidad y tipo de elementos que componen el

puente. No toma en cuenta variables como las características de tráfico, materialidades de elementos y las dimensiones de los puentes al calificar.

Las bases de datos de tráfico, catastro e inspección de puentes combinan variables numéricas, categóricas ordinales y categóricas nominales. Las herramientas de *Machine Learning* permiten analizar estas variables con la búsqueda de patrones en problemas de predicción, clasificación y agrupación de datos, lo cual no es posible con métodos estadísticos tradicionales.

3 HERRAMIENTAS DE MACHINE LEARNING PARA LA MEDICIÓN DE INFLUENCIA DE VARIABLES EXPLICATIVAS

3.1 Introducción

En este capítulo se describen distintas técnicas de *Machine Learning* (ML), las ventajas y desventajas que tienen, una breve explicación de cómo funcionan y las aplicaciones que han tenido en otros estudios de relevancia de atributos en distintos activos.

3.2 Clasificación de técnicas de *Machine Learning*

Las técnicas de ML se clasifican en aprendizaje supervisado, no supervisado o reforzado según las características de los datos de entrenamiento. En el entrenamiento supervisado la variable objetivo está registrada, se usa en tareas de regresión y clasificación. En el aprendizaje no supervisado no se presenta una variable objetivo sino un conjunto de variables independientes entre sí, se usa para problemas de agrupación de datos o de reducción dimensional de estos. El aprendizaje reforzado utiliza condiciones iniciales y retroalimenta iterativamente un modelo (Omar et al., 2022). En la Figura 3.1 se clasifican algunas técnicas de ML por tipo de aprendizaje.

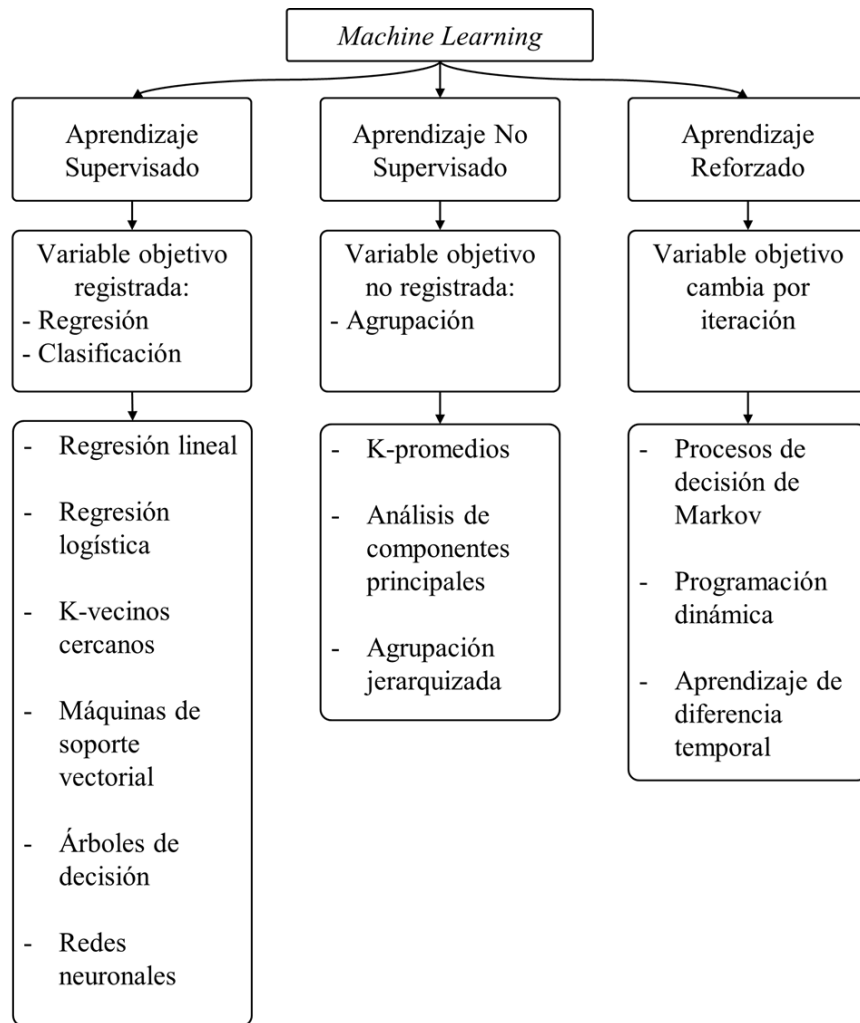


Figura 3.1 Clasificación de técnicas de Machine Learning (adaptado de Omar et al., 2022)

En la Tabla 3.1 se aprecia una descripción de distintas técnicas de aprendizaje supervisado, con una descripción de su funcionamiento general, sus usos, ventajas y desventajas.

Tabla 3.1 Catálogo de técnicas de Machine Learning

Técnica	Descripción
Regresión Lineal y Regresión Lineal Múltiple	Esta técnica ajusta una función lineal a un conjunto de datos numéricos usando métodos algebraicos, es usada para generar predicciones de una variable numérica. Presenta pocos parámetros de ajuste en comparación con otras técnicas y puede incluir distintas variables explicativas, numéricas o

	<p>categorías, creando una combinación lineal de los atributos, pero pierde interpretabilidad cuando se incluyen muchas variables ya que se termina ajustando a la base de datos de entrenamiento y no una tendencia general.</p>
<p>Regresión Logística o Regresión Logit - Multinomial</p>	<p>Ajusta una función sigmoide a un conjunto de datos de variable categórica binaria. La función resultante presenta valores entre 0 y 1 denotando la probabilidad de pertenecer a una categoría. Al igual que la regresión lineal pierde interpretabilidad cuando se incluye muchas variables.</p>
<p>k-Vecinos Cercanos</p>	<p>Clasifica datos con la misma clasificación de la mayoría de los k- registros más cercanos según su distancia en un espacio n-dimensional, presenta el parámetro de ajuste “k” donde se decide cuántos vecinos serán contados en la clasificación. No existe un proceso de ajuste o decisión que ayude a ver la influencia de distintas variables. La técnica es sensible en datos de entrenamiento desbalanceados y propensa a sobreajuste (Malekloo et al., 2022), pero efectiva en grandes bases de datos (Sun et al., 2020).</p>
<p>Máquinas de Soporte Vectorial</p>	<p>Lleva variables numéricas a dimensiones de mayor orden, donde genera hiperplanos que se ajusten a la separación de los datos en la nueva distribución espacial, clasificándolos como parte de un conjunto u otro.</p> <p>Es efectivo en espacios de varias dimensiones (muchas variables) pero es poco interpretable. Además, necesita como parámetro de ajuste una función kernel (núcleo) utilizada en la generación del hiperplano que es difícil de ajustar para distintas bases de datos (Malekloo et al., 2022).</p>
<p>Redes Neuronales</p>	<p>Transforma las variables de entrada a través de nodos interconectados que contienen funciones de activación (funciones que transforman la variable de entrada), el resultado de la transformación se ajusta a la variable objetivo que puede ser categórica o numérica. La transformación de variables dificulta la interpretabilidad del modelo.</p>

	<p>Existe una diversidad de versiones de redes neuronales que se especializan en distintas tareas, funcionan a través de algoritmos complejos que son difíciles de ajustar sin un estudio extenso. La técnica es adecuada para problemas sin una relación aparente a través de funciones, pero propensa a sobreajuste (Sun et al., 2020).</p>
<p>Árboles de Decisión</p>	<p>Genera un conjunto ramificado de nodos de decisión binaria. Combina el uso de variables independientes numéricas y categóricas de un registro para generar una decisión respecto a la variable de salida (también numérica o categórica) en base a la pertenencia a una clase o un umbral numérico.</p> <p>Los parámetros de ajuste son fáciles de configurar ya que se enfocan en la cantidad de niveles de nodos presentes en la ramificación o la cantidad de registros que limita que un nodo se subdivide en un nodo siguiente. Se adapta a problemas donde no existe relación directa entre variables explicativas y variable objetivo. Fácil de interpretar, poco sensible a datos atípicos, propenso a sobreajuste (Sun et al., 2020).</p>
<p>Random Forest</p>	<p>Conjunto masivo de árboles de decisión generados a partir de segmentos aleatorios de los datos. El conjunto de árboles vota para decidir la clasificación de la variable objetivo o predecir un valor de esta. Absorbe ventajas de árboles de decisión respecto a los parámetros presentes en la técnica y a la adaptación a problemas complejos. Es eficaz en grandes bases de datos, disminuye el ser propenso a sobreajuste utilizando distintos segmentos de las bases de datos.</p>
<p>Adaptative Boosting</p>	<p>Conjunto de árboles de un solo nodo de decisión, que son ponderados a través de validación de las predicciones para cuantificar si las variables explicativas aportan o no a la decisión. Presenta la habilidad para reducir el sesgo al momento de elegir una técnica de clasificación, pero es una técnica diseñada para complementar otras herramientas de ML. Es sensible a datos atípicos (Sun et al., 2020).</p>

Extreme Gradient Boosting	Conjunto de árboles de decisión agregados gradualmente que ajustan los residuos o errores de clasificación del árbol anterior. Es una técnica diseñada con distintos parámetros de regularización, relleno de datos vacíos, segmentación de datos y ventajas computacionales que optimizan su desempeño. Absorbe ventajas de árboles de decisión ya que sigue ajustándose a problemas complejos incluyendo variables independientes numéricas y categóricas simultáneamente, aunque presenta la ventaja de lidiar con registros con variables vacías.
---------------------------	---

3.3 Aplicaciones de ML en medición de influencia de atributos

En distintos modelos ingenieriles las variables consideradas relevantes son generalmente seleccionadas según la experiencia ingenieril (Zhu y Wang, 2021) mas no por la cuantificación de la influencia de estas variables en un modelo. La calibración de modelos es sensible a incluir variables explicativas que no aporten realmente a la predicción o clasificación de un fenómeno, generando reducciones de exactitud y pérdida de interpretabilidad de los modelos. Los estudios de medición de influencia de las variables explicativas son usados para reducir la cantidad de variables en un modelo sin perder su capacidad de predicción o clasificación.

Zhu y Wang (2021) elaboraron un ranking de influencia de 25 variables explicativas de condición de puentes usando la técnica ReliefF, para generar una selección más eficiente de variables a monitorear. Utilizaron una base de datos de 23104 registros de puentes de Texas, EEUU, entre 1992 y 2019. ReliefF usa la técnica k-Vecinos Cercanos para encontrar registros vecinos a un registro fijo, compara el contenido de las variables entre registros y aumenta la ponderación de influencia de las variables que diferencian a registros con distinta clasificación que el registro fijo. Los autores organizaron los ponderadores de cada variable en un diagrama de Pareto y se determinó que las variables que contengan el 90% de la acumulación de la calificación son las más relevantes.

Omar et al. (2022) investigaron la importancia de cuatro variables explicativas de la propagación de grietas en elementos de una aleación metálica. Se ajustó un modelo de regresión lineal regularizado para obtener los coeficientes para cada una de las variables explicativas, luego con los coeficientes normalizados entre sí de cada variable comparan su importancia relativa en la explicación del modelo.

Los autores muestran cómo se puede determinar la influencia de variables en la descripción de un fenómeno a través de modelos que no apunten directamente a la ponderación de influencia de variables.

El Moutassir et al. (2016) determinaron las variables influyentes en la detección de corrosión en tuberías. Para esto usaron la técnica de clasificación de k-Vecinos Cercanos para clasificar tuberías como corroídas o no en conjunto con *Sequential Forward Feature Selection* (SFFS). La técnica SFFS agrega secuencialmente las variables que minimicen el error de clasificación, el proceso continúa hasta que agregar más variables no aumente significativamente el rendimiento del modelo. Obtuvieron las tres variables explicativas más influyentes de un conjunto de once variables medidas.

Breiman (2001) introduce la Importancia de Atributos por Permutación como método de medición de influencia de variables. Esta es una técnica que necesita modelos de ML supervisados ya entrenados en clasificación o regresión, funcionando de forma complementaria. Los modelos requieren medidas de rendimiento de clasificación como Sensibilidad (que se define como la proporción de clasificaciones correctas sobre las instancias reales) o rendimiento de predicción como el Coeficiente de determinación o el Error cuadrático medio. La permutación intercambia aleatoriamente los valores de una variable explicativa entre los registros, generando un nuevo conjunto de registros para testear el modelo. Se mide nuevamente el rendimiento de clasificación o predicción del modelo con los nuevos registros modificados como datos de entrada. Si al intercambiar los valores en la variable explicativa entre los registros se reduce el rendimiento, entonces la variable explicativa realmente afecta a la variable objetivo. En cambio, si al intercambiar los valores en la variable explicativa no hay una disminución en rendimiento entonces la variable explicativa no afecta a la variable objetivo.

3.4 Conclusiones

Existen diversas técnicas de *Machine Learning*, utilizadas en reconocimiento de patrones de clasificación, predicción o agrupación de datos y divididas en aprendizaje supervisado (variable objetivo registrada), aprendizaje no supervisado (sin variable objetivo) y aprendizaje reforzado (variable objetivo retroalimentada en iteraciones). La calificación de condición de puentes registrada en las bases de datos de inspección de puentes es considerada la variable objetivo, por lo que las técnicas de ML utilizadas para el análisis serán de aprendizaje supervisado.

La aplicación de técnicas de ML para la búsqueda de patrones depende de la tarea modelada (predicción, clasificación o agrupación de datos) y de los tipos de datos con los que se entrenen (categóricos, numéricos o ambas). Las técnicas presentan distintos niveles de complejidad en sus parámetros de ajuste. Por estas razones, es necesario tener distintos criterios de selección de técnicas de ML para la búsqueda de patrones, teniendo en cuenta las características de las variables independientes y dependiente a modelar.

Es necesario usar técnicas complementarias para la medición de influencia de variables explicativas en modelos de ML, ya que es posible que las variables no puedan interpretarse independientemente entre sí en los resultados de un modelo. Estas técnicas complementarias pueden funcionar en conjunto a una técnica principal de clasificación, predicción o agrupación de datos que debe adaptarse a las características del problema modelado.

Seleccionar las variables explicativas más influyentes en un modelo y descartar las variables menos influyentes permite aumentar la exactitud de los modelos y su interpretabilidad, además de reducir el sesgo de elección de variables según experiencia ingenieril y no su verdadero impacto en los resultados.

4 METODOLOGÍA DE ANÁLISIS CON HERRAMIENTAS DE MACHINE LEARNING

4.1 Introducción

En este capítulo se detalla el procedimiento que permite generar los rankings de influencia relativa de las variables, pasando por la configuración de la base de datos necesaria para implementar las técnicas de ML, los criterios de selección de estas técnicas y su procedimiento de modelación en software.

4.2 Diagrama de relación de actividades por objetivo específico

En la Figura 4.1 se presenta un diagrama de las actividades realizadas en la memoria, conectado distintas actividades relevantes a través de flechas indicando su secuencia. En cada columna, se muestra una actividad en la parte superior, seguido de una secuencia de sub – actividades hacia abajo y su resultado correspondiente.

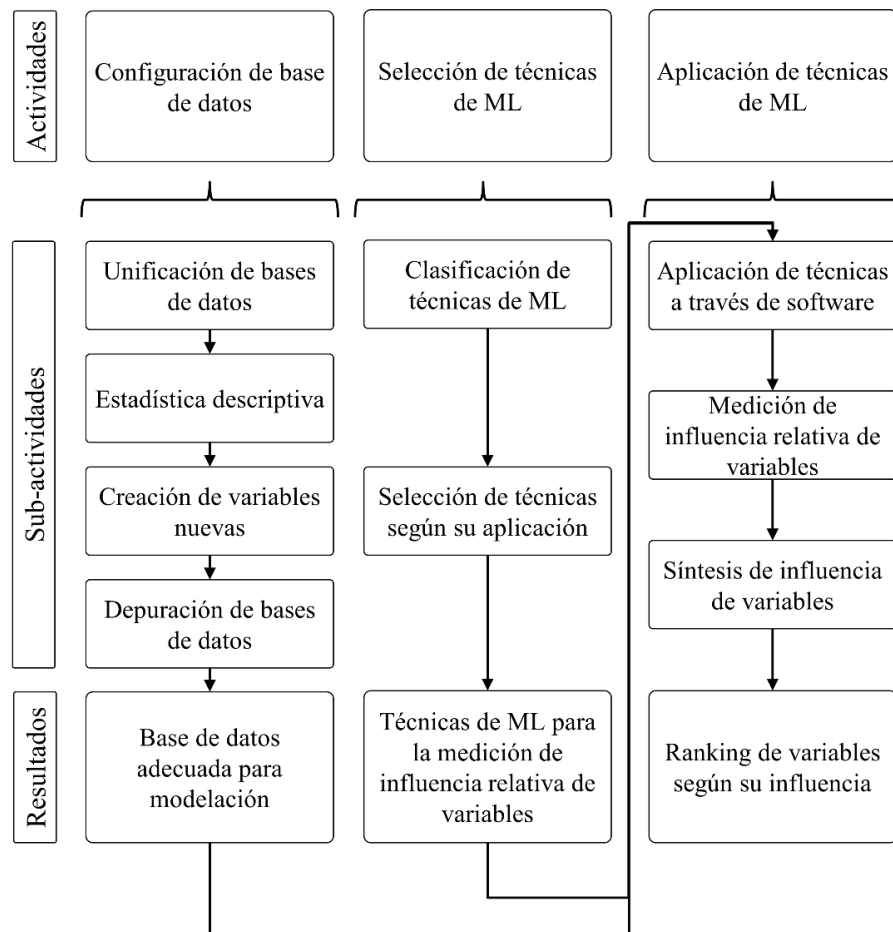


Figura 4.1 Diagrama de relación de actividades

4.3 Configuración de base de datos

Se dispone de dos bases de datos por región. Una de tráfico y otra de catastro e inspección de puentes. Se identifica el contenido de variables en cada base de datos además de su cantidad de registros. Las bases de datos de tráfico son unificadas con las bases de datos de inventario e inspección de puentes por región, coincidiendo la ubicación del puente (registrada en la base de datos de inventario) con la ubicación de un tramo de ruta (registrada en la base de datos de tráfico).

Se utiliza estadística descriptiva para ver la relación entre variables independientes y variable dependiente. Se describe la calificación de condición por tipología estructural con su promedio, rango, y desviación estándar. Se crean diagramas de caja de la calificación de condición por materialidad de elementos y extensión o gravedad de los deterioros. Por último, se genera un ranking de la cantidad de deterioros recurrentes en cada región.

Se crean nuevas variables que resuman la información de variables semejantes o que relacionen distintas variables existentes entre sí. Para esto se almacena en variables nuevas la suma de variables semejantes o el producto de variables distintas que se quiera relacionar respectivamente. Por ejemplo, se suma la cantidad de TMDA de autos y camionetas en la variable TMDA de vehículos livianos y se divide la cantidad de deterioros entre la longitud total del puente en la variable deterioros por longitud de puente.

Para la depuración de las bases de datos unificadas se cuantifica la cantidad de datos sin información por variable, excluyendo las variables que presentaron más de un 50% de vacíos. Se excluyen del análisis los registros de puentes que no hayan sido inspeccionados (es decir, sin calificación de condición y/o sin deterioros registrados), registros que no cuenten con las dimensiones del puente y por último registros que no cuantifiquen el tráfico presente en el puente.

4.4 Selección de técnicas de ML

Las técnicas de ML utilizadas para el análisis deben cumplir con los siguientes criterios de selección:

1. Contexto de aprendizaje supervisado, incluyendo variables numéricas y categóricas para la medición de influencia simultáneamente. Debido a la necesidad de medir la influencia de

variables de distinto tipo, además, es favorable mantener la escala de medición de influencia en una misma técnica y no separar los tipos de variable.

2. Capacidad modelar problemas con falta de relación directa entre variables dependientes e independientes, debido a diversidad y cantidad de variables explicativas a medir.
3. Disponibilidad de software accesible, evitando complejos parámetros de modelación. Este criterio es necesario debido a que la modelación de los algoritmos de las técnicas es secundaria a la memoria y no parte de los objetivos a desarrollar.
4. Capacidad de trabajo con datos no homogéneos entre sí. Este criterio es necesario debido a que los registros presentan vacíos en distintas variables entre sí. Es el último criterio ya que es posible solventarlo en una etapa de depuración más específica de ser necesario.

Las técnicas seleccionadas usan complementariamente la Importancia de Atributos por Permutación para la medición de influencia de variables explicativas.

4.5 Aplicación de técnicas de ML

La aplicación de técnicas de ML se modela por separado para las bases de datos unificadas por región. La modelación incluye la segmentación de datos por tipo de tarea, el entrenamiento de las técnicas de ML en tareas de regresión y clasificación por separado y la medición de la influencia de variables explicativas. Para evitar el sobreajuste se utiliza validación cruzada de cinco segmentos en el entrenamiento de las técnicas. La validación cruzada segmenta el conjunto de datos en “ k ” conjuntos, el modelo se entrena excluyendo uno de los conjuntos para usarlo como datos de validación y se repite el proceso hasta que cada uno de los k conjuntos haya sido excluido del proceso de entrenamiento y utilizado como conjunto de validación.

Se mide el rendimiento de predicción o clasificación de las técnicas de ML modeladas con los parámetros de Coeficiente de determinación y Sensibilidad respectivamente, para luego aplicar la técnica de Importancia de Atributos por Permutación para la medición de influencia de variables en las técnicas entrenadas.

Se mide la influencia de variables explicativas en dos etapas. En la primera etapa se incluyen todas las variables explicativas en cada técnica por separado. Se descartan del ranking de influencia las variables que tengan un puntaje de influencia de un orden de magnitud menor al de la variable más influyente. En la segunda etapa se filtra el conjunto de variables, utilizando las variables que hayan

sido incluidas como influyentes por al menos una técnica y se mide la influencia de las variables filtradas exclusivamente. Los puntajes de influencia pueden ser desequilibrados entre distintas técnicas, por lo que los resultados se normalizan entre el mayor y el menor puntaje por técnica, reflejando la posición relativa en un ranking de influencia. Por último, se promedia la influencia normalizada de cada variable.

4.6 Conclusiones

El análisis descriptivo de las variables permite entender la relación entre variables independientes y dependiente. Esto orienta la selección de técnicas de ML a implementar. La creación de variables nuevas posibilita medir la influencia de las relaciones entre variables que no estén explícitamente registradas en el sistema de información pero que pueden ser derivadas de este.

Los criterios de selección de técnicas de ML permiten reducir la variedad de técnicas disponibles adecuándose al contexto de las bases de datos utilizadas en el análisis, incluyendo la Importancia de Atributos por Permutación en la medición de influencia de variables explicativas.

La técnica de medición de influencia puede cuantificar tanto variables numéricas como categóricas, adecuándose al problema de la memoria. Normalizar el puntaje de medición de la influencia de las variables permite compararlas a través de técnicas de distinto funcionamiento, pudiendo interpretar en una sola medición final promedio la influencia de las variables explicativas.

5 CONFIGURACIÓN DE BASE DE DATOS

5.1 Introducción

Como parte de la configuración o preprocesamiento de la base de datos a utilizar en la identificación de influencia de variables se analizó el contenido de las variables, se explica el proceso de unificación de las bases de datos, la creación de variables extra y la depuración la base de datos para conservar solo los registros útiles para la aplicación de las técnicas de ML seleccionadas.

5.2 Fusión de base de datos

La memoria cuenta con bases de datos de dos de las 16 regiones del país, una región ubicada en la zona central y otra de una región ubicada en la zona sur. Para cada región, se cuenta con dos bases de datos, una de tráfico de vehículos y otra de catastro e inspección de puentes.

Las bases de datos de tráfico contienen cuatro variables de identificación y nueve variables cuantitativas. Las variables de identificación son: código de ruta, kilómetro de inicio de tramo, kilómetro de fin de tramo y numeración de tramo. Las variables cuantitativas son: tráfico medio diario anual (TMDA), proporción de autos, proporción de camionetas, proporción de camiones de dos ejes, proporción de camiones de más de dos ejes, proporción de remolques, proporción de semi-remolques, proporción de buses y velocidad del tramo. Se asocia un total de 13 variables a cada tramo registrado, con 1267 tramos en la base de datos de la región de la zona centro y 306 tramos en la base de datos de la región de la zona sur.

Las bases de datos de catastro e inspección de puentes contienen 10 variables de catastro administrativo (relativas a la identificación del puente), 19 variables de catastro técnico (relativas a las dimensiones y materialidad de los elementos del puente) y 19 variables de inspección (relativas a la condición del puente). En la Tabla 5.1 se muestran las variables de catastro e inspección de puentes mencionadas. A cada deterioro registrado en las bases de datos se asocia un total de 48 variables, con 4041 deterioros registrados en la base de datos de la región de la zona centro y 2999 deterioros registrados en la base de datos de la región de la zona sur. En la [A2.1.1 del Anexo 2.1](#) se detalló el nombre, contenido y descripción de cada variable.

Tabla 5.1 Variables en bases de datos de catastro e inspección de puentes presentes

Variables de catastro administrativo	Variables de catastro técnico	Variables de inspección
Nombre Puente Código Puente Código I3MOP Región Provincia Comuna Tuición Código de Ruta Clase de Ruta km	Tipología General Tipología Estructural Longitud total (m) Ancho total (m) Número de tramos Año de construcción Número de Cepas Ancho de Calzada (m) Ancho de Pasillo (m) Espesor de Losa (m) Número de Vigas Material de cepas Tipo de Fundación de cepas Material de estribos Tipo de Fundación estribos Material de fundación de estribos Carpeta Vigas Losa	Fecha inspección principal Calificación automática Calificación final Número total de deterioros Índice de Estructural Índice de Seguridad Vial Índice de Contención Sistema de Contención Índice de Contención Barandilla Índice de Durabilidad Índice de Daños Menores Deterioro Causa Gravedad Extensión Medición Unidad medición Calificación automática Unidad Código

Para unificar las bases de datos por región se buscó la coincidencia entre la ubicación del puente (en la base de datos de catastro e inspección) y la ubicación de un tramo de ruta (en la base de datos de tráfico). Identificando el tramo de ruta al que pertenece un puente se agregó la información del tramo al conjunto de variables asociadas a cada deterioro de puente en las bases de datos de catastro e inspección. Las bases de datos unificadas tienen 58 variables de 4041 deterioros registrados en la base de datos de la región de la zona centro y 2999 deterioros registrados en la base de datos de la región de la zona sur.

5.3 Estadística descriptiva de variables

La calificación de condición fue descrita por segmento de Tipología estructural, Gravedad y Extensión de deterioro y materialidad de distintos elementos en distintas tablas y diagramas de caja para describir el contenido de la base de datos. En la Tabla 5.2 y Tabla 5.3 se aprecian los segmentos de Tipología estructural por región, la cantidad de registros asociados a cada tipología, el porcentaje de registros de esta tipología respecto al total de registros de la región, la calificación final mínima, máxima, promedio de la tipología y la desviación estándar de esta por región.

Tabla 5.2 Estadística descriptiva por Tipología Estructural – Región zona centro

Tipología estructural	Cantidad de registros	Porcentaje de registros respecto a total de registros en la región (%)	Calificación Mínima	Calificación Máxima	Calificación Promedio	Desviación Estándar de Calificación
Arco	66	2	50	100	57	15
Bóveda	2	0	71	71	71	0
Convencional cajón	409	10	42	100	67	10
Convencional celosía	131	3	45	65	55	8
Convencional losa	351	9	45	85	73	9
Convencional vigas	2930	72	25	100	60	14
Marco / Cajón / Pórtico	30	1	45	70	61	8
Otros	16	0	41	83	54	19
(en blanco)	106	3	40	100	69	12
Total general	4041	100	25	100	62	14

Para la región de la Zona Centro, la calificación de condición en la región varía de 25 a 100 en los 4041 registros de deterioro, con una calificación promedio de 62 y una desviación estándar de 14 (23% del promedio). Se muestra la dispersión de las calificaciones por tipología en diagramas de caja en la Figura 5.1. La mayoría de los puentes son de tipo Convencional de Vigas (72%), donde además se presenta el rango más amplio de calificación y la menor calificación mínima. Las otras tipologías no superan el 10% de los registros. La tipología de bóveda presenta dos deterioros en el mismo puente por lo que no tiene variación. El 3% de los registros no presentan una tipología definida (Otros o sin tipología registrada).

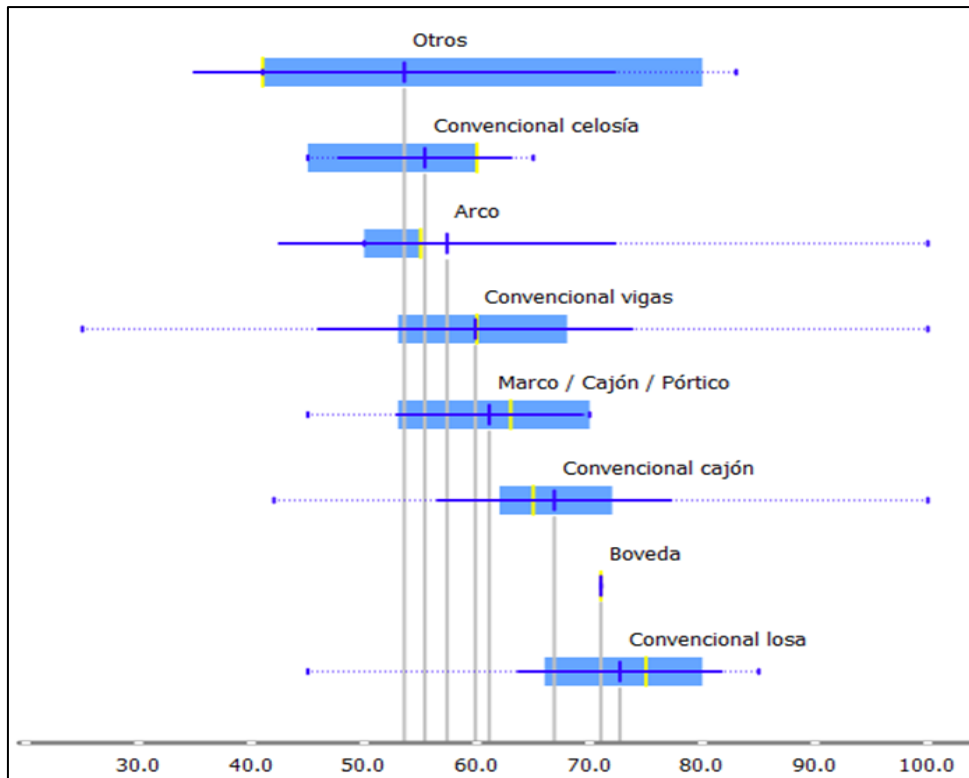


Figura 5.1: Distribución de calificación de condición por Tipología Estructural – Región zona centro

La Figura 5.1 muestra en el eje horizontal la calificación de condición de puentes. En los diagramas de caja, las líneas verticales gris indican el promedio, las líneas verticales amarillas indican la mediana, la línea azul horizontal indica una desviación estándar más y menos desde el promedio y los bigotes del diagrama representan los datos extremos. Los diagramas de caja se superponen entre sí entre las calificaciones 45 y 85 aproximadamente, no hay una distinción específica de una tipología sobre las otras.

Tabla 5.3 Estadística descriptiva por Tipología Estructural – Región zona sur

Tipología Estructural	Cantidad de registros	Porcentaje de registros respecto a total de registros en la región (%)	Calificación Mínima	Calificación Máxima	Calificación Promedio	Desviación Estándar de Calificación
Arco	53	2	65	100	72	6
Colgante	188	6	39	100	57	17
Convencional cajón	282	9	38	100	64	12
Convencional celosía	71	2	50	100	60	12
Convencional losa	11	0	75	75	75	0
Convencional vigas	2326	78	19	100	64	13
Otros (en blanco)	51	2	61	100	71	7
	17	0	70	100	83	14
Total general	2999	100.0	19	100	64	13

Para la región de la Zona Sur, la calificación de condición en la región varía de 19 a 100 en los 2999 registros, con una calificación promedio de 64 y una desviación estándar de 13 (20% del promedio). Se muestra la dispersión de los datos en diagramas de caja en la Figura 5.2. La mayoría de los puentes son de tipo Convencional de Vigas (del 78 %), donde además se presenta el rango más amplio de calificación y la menor calificación mínima. La tipología de bóveda presenta once deterioros en el mismo puente por lo que no tiene variación. El 2% de los registros no presentan una tipología definida (Otros o sin tipología registrada).

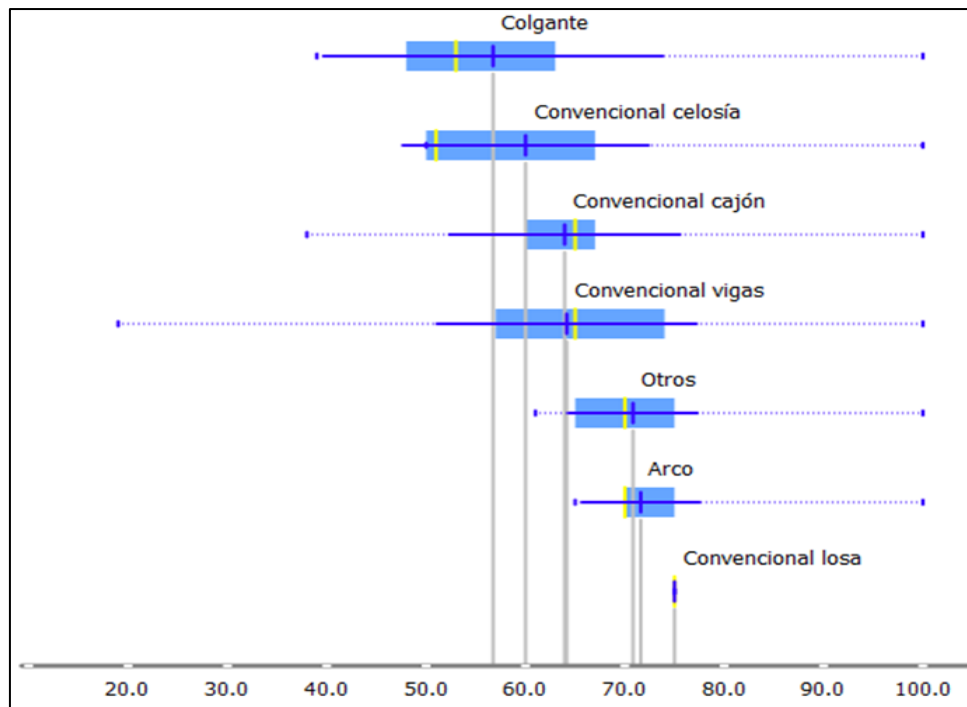


Figura 5.2: Distribución de calificación de condición por Tipología Estructural – Región zona sur

La Figura 5.2 muestra diagramas de caja que se superponen entre sí, entre las calificaciones 50 y 100 aproximadamente, a excepción de la tipología convencional de losa que presenta registros de un mismo puente no hay una distinción específica de una tipología sobre las otras.

En las Figuras Figura 5.3 a la Figura 5.8 se aprecia la distribución de calificaciones finales en diagramas de caja según materialidad de cepas, vigas y losas por región.

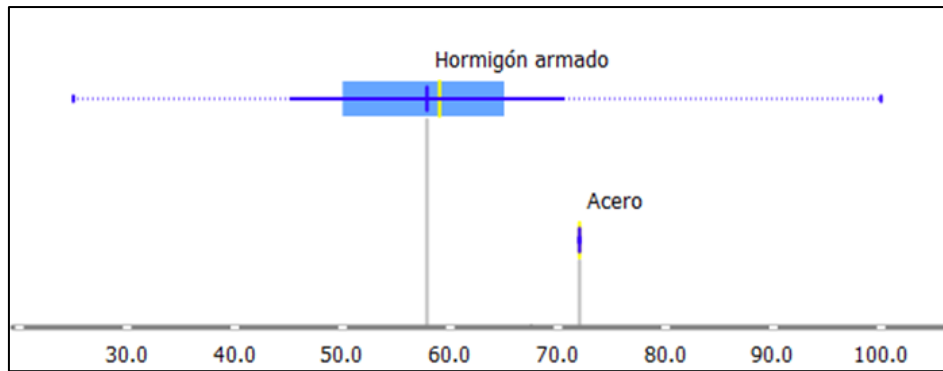


Figura 5.3: Distribución de calificación de condición por material de cepas – Región zona centro

En la Figura 5.3 la materialidad de cepas de acero presenta un solo puente con solo una calificación de condición por lo que no presenta dispersión.

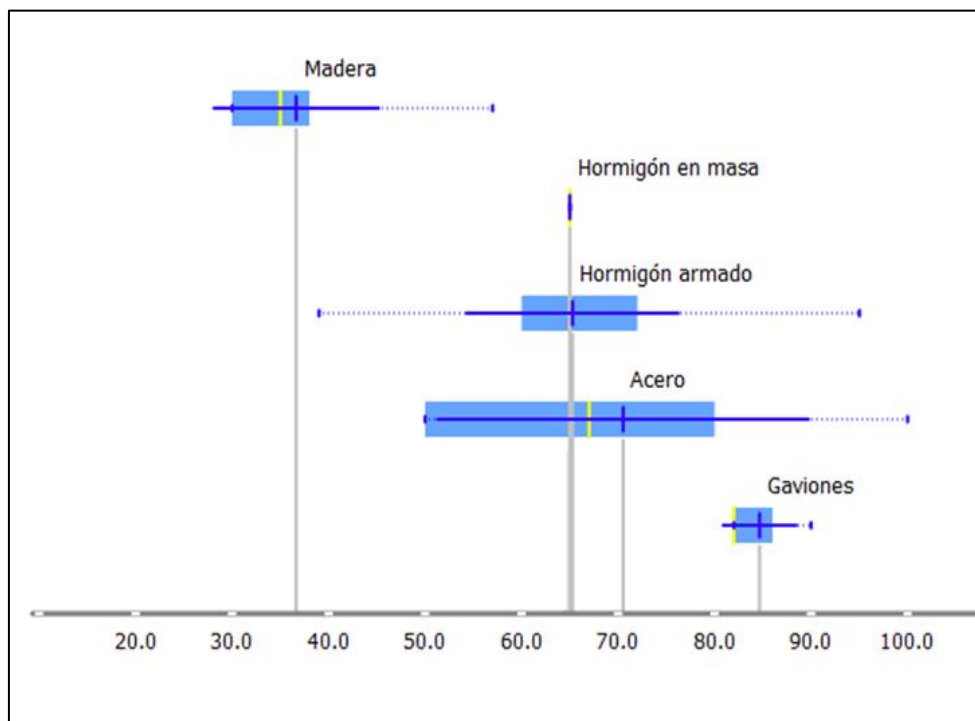


Figura 5.4: Distribución de calificación de condición por material de cepas – Región zona sur

En la Figura 5.4 se muestra que el diagrama de caja de las cepas de madera se aleja significativamente de las otras materialidades y que la materialidad de hormigón en masa no presenta dispersión (al igual que en casos anteriores por ser registros de un solo puente con una sola calificación de condición).

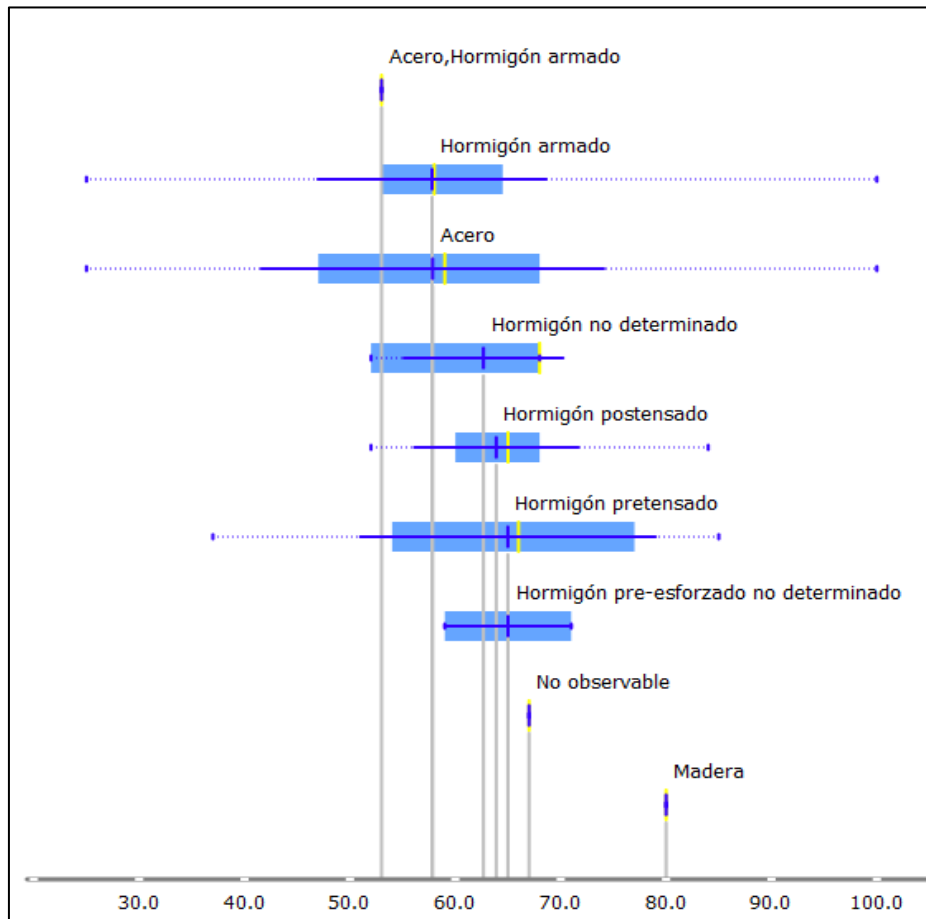


Figura 5.5: Distribución de calificación de condición por material de vigas – Región zona centro

En la Figura 5.5 se muestra que los diagramas de caja de hormigón armado y de acero abarcan casi todas las calificaciones desde 25 aproximadamente a 100. Además, casi todos los diagramas de caja se superponen entre sí entre las calificaciones entre 50 y 90 aproximadamente para casi todas las materialidades, a excepción de aquellas sin dispersión por tener registros de un solo puente por lo tanto una sola calificación. La materialidad de hormigón pre – esforzado no determinado presenta un rango menor a las otras materialidades indicando una distinción del resto.

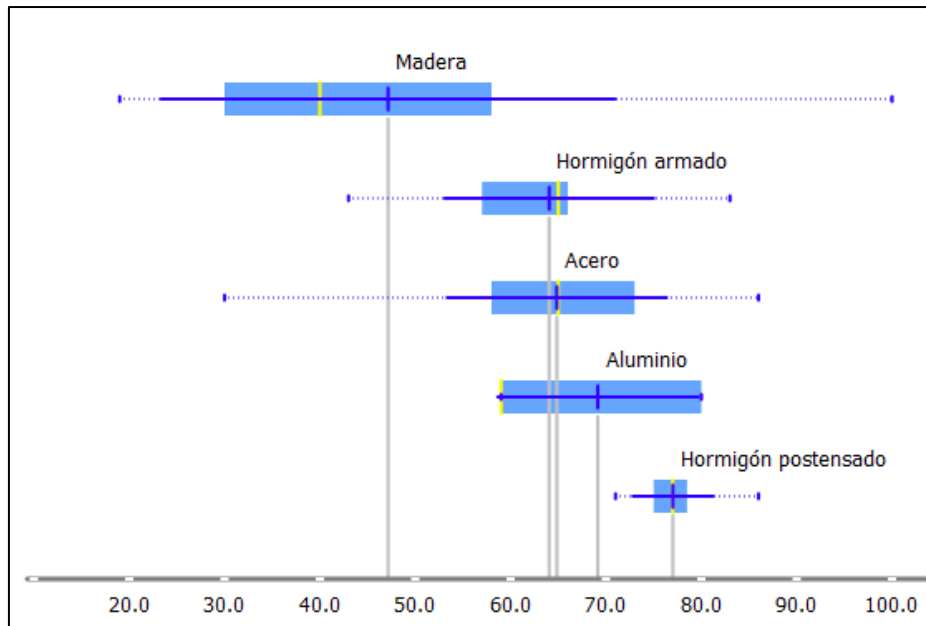


Figura 5.6: Distribución de calificación de condición por material de vigas – Región zona sur

En la Figura 5.6 se muestra una cantidad menor de materialidades respecto a aquellas de la región de la zona centro. La materialidad de madera abarca el mayor rango de calificaciones presente, entre 20 y 100 aproximadamente, con la menor calificación promedio de todas las materialidades. Las materialidades de aluminio y de hormigón postensado presenta rangos más pequeños que las otras, distinguiéndose de las demás.

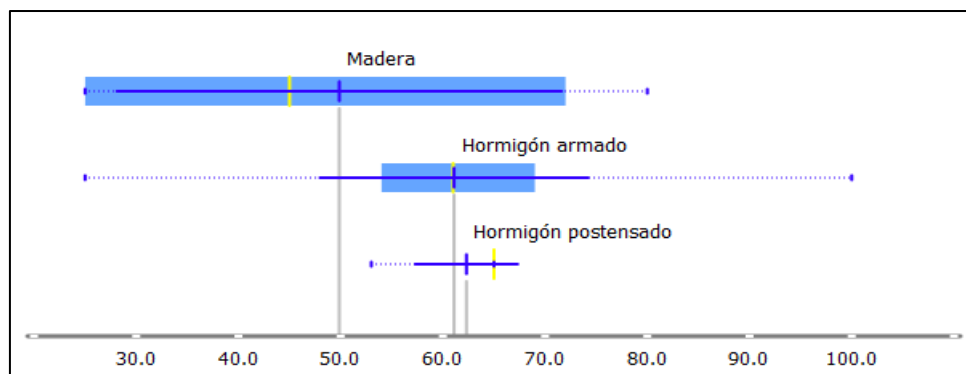


Figura 5.7: Distribución de calificación de condición por material de losas – Región zona centro

En la Figura 5.7 se muestra que el promedio de la materialidad de madera es el menor pero que abarca un rango menor de calificaciones que la materialidad de hormigón armado.

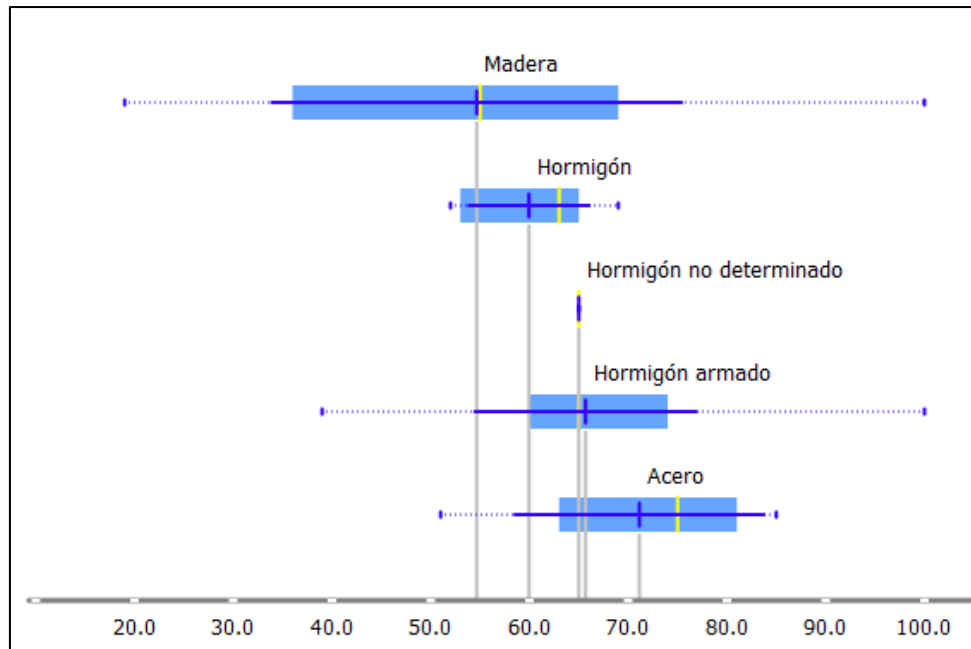


Figura 5.8: Distribución de calificación de condición por material de losas – Región zona sur

En la Figura 5.8 se muestra que nuevamente la materialidad de madera es la de menor calificación de condición promedio, abarca el mayor rango de calificaciones. La materialidad de hormigón presenta un rango menor que el resto distinguiéndose. Las materialidades de hormigón armado y de acero se superponen entre sí en conjunto con la de madera con amplios rangos.

En la Tabla 5.4 se muestra un ranking de los 10 tipos de deterioro más recurrentes por región según el porcentaje respecto del total de registros de cada región. Los deterioros más recurrentes (Vegetación y Eflorescencias) no afectan a nivel estructural y requieren una intervención mínima por lo que tienen una incidencia menor en la calificación final (refiriéndose al coeficiente de ponderación de Vegetación y Eflorescencias en la Tabla A2.2.2).

Tabla 5.4 Ranking de Tipos de deterioro recurrentes

Región	Tipo de Deterioro	Cantidad de registros por tipo de deterioro	Porcentaje respecto a total de registros en la región (%)
Zona centro	Eflorescencias	385	10
	Desprendimientos de hormigón con armadura vista	339	8
	Grietas (>5 mm)	275	7
	Alteración superficial	274	7
	Grafiti	260	6
	Humedades, filtraciones	258	6
	Desprendimientos de hormigón sin armadura vista	250	6
	Fisuras	237	6
	Depositación	211	5
	Corrosión	173	4
Zona sur	Vegetación	471	16
	Depositación	400	13
	Humedades, filtraciones	351	12
	Pérdida de tratamiento protector	262	9
	Pátina, mancha de óxido	157	5
	Eflorescencias	130	4
	Deformación	111	4
	Desprendimientos de hormigón sin armadura vista	99	3
	Pérdida de pieza	93	3
	(en blanco)	90	3

En la Tabla 5.5 se muestran los porcentajes del total de registros de las categorías de Gravedad y categorías de Extensión. Las tablas se leen intersecctando la categoría de gravedad de las filas y la categoría de extensión de las columnas.

Tabla 5.5 Distribución de Gravedad y Extensión en deterioros

Región	Extensión	Extensión	Extensión	Extensión	Extensión muy	Total general (%)
	Gravedad	pequeña (%)	media (%)	elevada (%)	elevada (%)	
Zona centro	Gravedad mínima (%)	67	9	4	7	87
	Gravedad media (%)	3	4	2	2	11
	Gravedad alta (%)	1	0	0	0	2
	Gravedad muy alta (%)	0	0	0	0	0
	Total general (%)	71	14	6	9	100
Zona sur	Gravedad mínima (%)	65	9	4	6	85
	Gravedad media (%)	3	4	2	2	11
	Gravedad alta (%)	1	0	0	0	2
	Gravedad muy alta (%)	0	0	0	0	0
	Total general (%)	69	14	6	8	97(*)

(*) La Región zona sur presenta un 3% de registros de deterioro donde no se encuentra clasificada ni su Gravedad ni su Extensión.

Para ambas regiones cerca del 65% de los deterioros son de Gravedad Mínima y Extensión Pequeña presentando una menor incidencia en la calificación final (detalle en cálculo de “Co” según la ecuación A2.2 del Anexo ANEXO 2.2 Cálculo de calificación de condición de puentes).

Se presentan los diagramas de caja con el desglose de las categorías independientes entre la Figura 5.9 y la Figura 5.12 para ambas regiones.

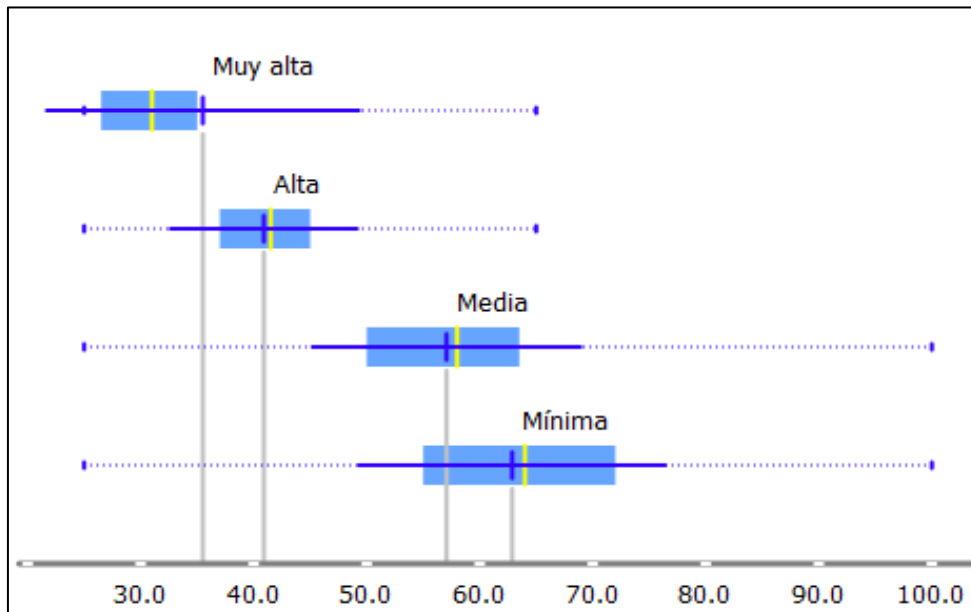


Figura 5.9 Calificación de condición por Gravedad – Región zona centro

En la Figura 5.9 se muestra que el rango de calificaciones de la gravedad muy alta es tan extenso como el de la gravedad alta, abarcando calificaciones entre 25 y 65 aproximadamente, mientras que los rangos de las gravedades media y mínima abarcan entre calificaciones entre 25 y 100.

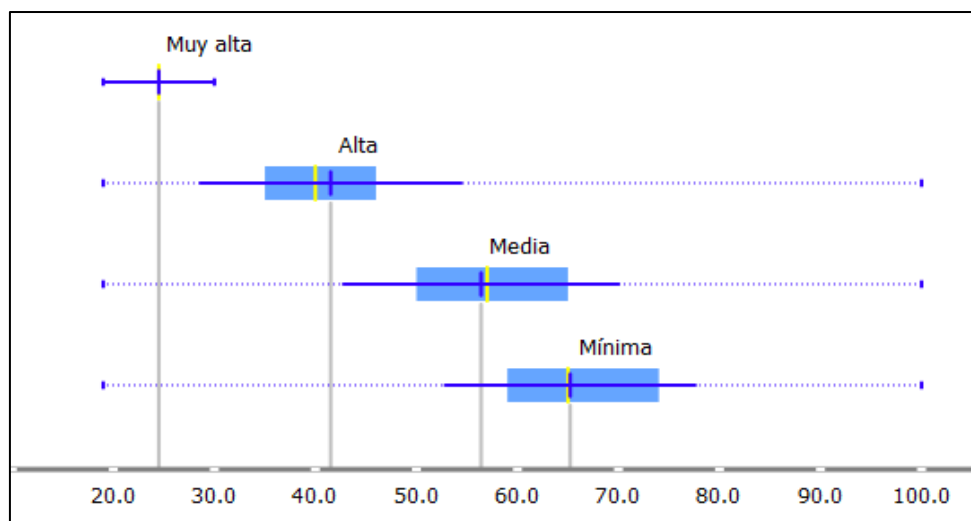


Figura 5.10 Calificación de condición por Gravedad – Región zona sur

En la Figura 5.10 se muestra que el rango de calificaciones de la gravedad muy es acotado y se distingue del resto de categorías, abarcando calificaciones entre 20 y 30 aproximadamente, mientras que los rangos de las otras categorías abarcan entre calificaciones entre 20 y 100.

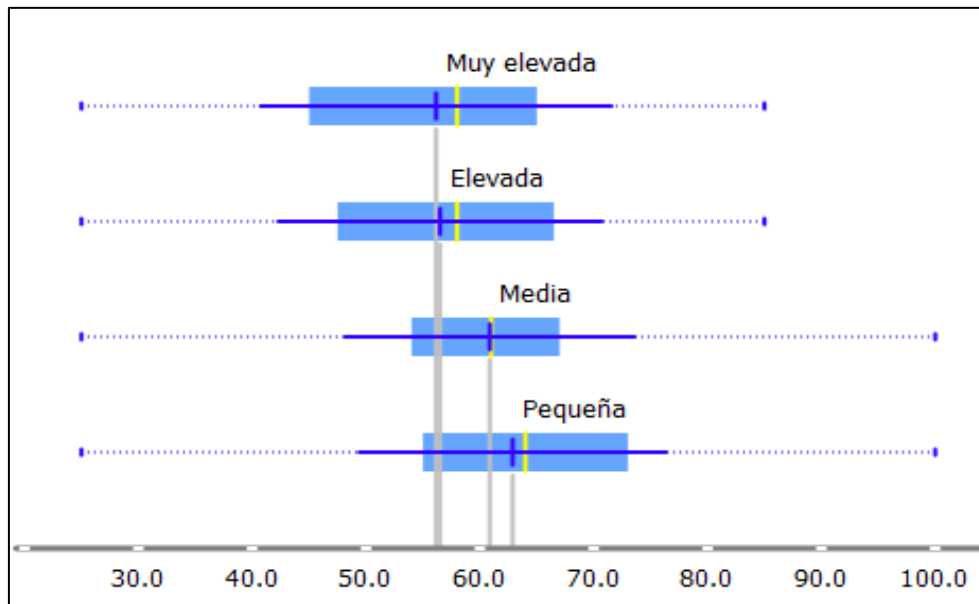


Figura 5.11 Calificación de condición por Extensión – Región zona centro

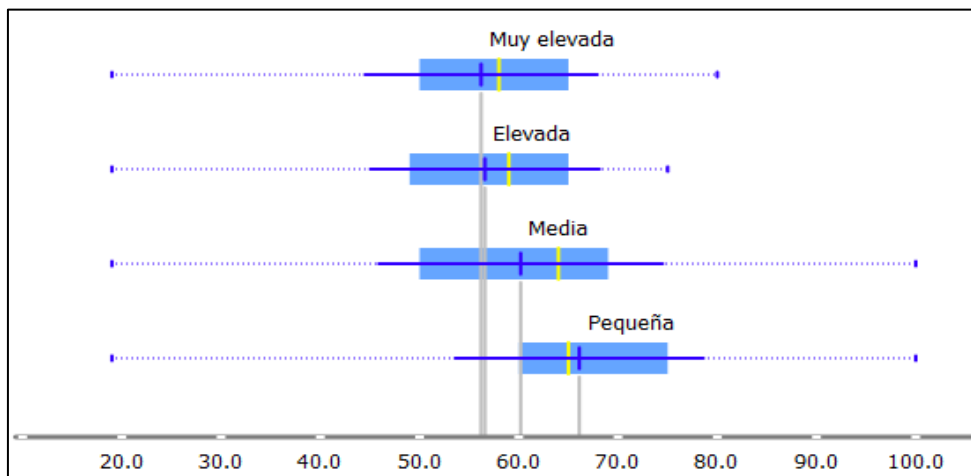


Figura 5.12 Calificación de condición por Extensión – Región zona sur

En la Figura 5.11 y la Figura 5.12 se muestra que los rangos de todas las categorías de extensión de deterioros se superponen entre sí, sin generar una distinción característica.

5.4 Variables incluidas en el análisis

A partir de las bases de datos originales y teniendo en cuenta el contenido de las variables presentes (58 en las bases de datos unificadas por región) se crearon variables que sinteticen distintas variables presentes en una o que generen nuevos parámetros considerados.

En la Tabla 5.6 se aprecian las variables creadas, su tipo y la descripción de estas. Las variables “cantidad de vehículos livianos”, “cantidad de vehículos pesados”, “proporción de proporción de vehículos livianos” y “proporción de vehículos pesados” fueron creadas debido a que sintetizan la información separada en ocho variables de tipos de vehículo en las bases de datos de tráfico. La variable “material menos resistente” fue creada porque se aprecia en la sección 5.3 la tendencia a que la madera tiene la menor calificación de condición promedio en todas las materialidades de elementos de la región de la zona sur, sintetizando la tendencia en una sola variable.

Las variables de deterioros por longitud de puente y deterioros por superficie de tablero fueron creadas para cuantificar la densidad espacial de deterioros presentes, relacionando cantidad de deterioros con las dimensiones del puente. La variable “calificación de condición” representa la calificación del puente en forma categórica. La variable “categoría de deterioro” sintetiza los 37 tipos de deterioro en cuatro categorías (se detalla qué tipo de deterioro corresponde a qué categoría de deterioro propuesta en la Tabla A5.3 del Anexo ANEXO 5.1 Estadística descriptiva de variables).

Tabla 5.6 Descripción de variables creadas

Nombre de variable	Tipo de variable	Descripción
TMDA de vehículos livianos	Numérica	TMDA del tramo de la ruta de vehículos livianos (Autos y Camionetas)
TMDA de vehículos pesados	Numérica	TMDA del tramo de la ruta de vehículos livianos (distinto de Autos y Camionetas)
Proporción de vehículos livianos	Numérica Porcentual	Proporción porcentual tráfico liviano de TMDA entre categorías de tráfico liviano y pesado
Proporción de vehículos pesados	Numérica Porcentual	Proporción porcentual tráfico pesado de TMDA entre categorías de tráfico liviano y pesado
Material menos resistente	Categórica	Material menos resistente presente en estructura por orden de prioridad: Madera < Aluminio < Hormigón < Acero
Calificación de condición	Categórica	Calificación de condición de puentes según categorías en Tabla 2.1 Etiquetas de categorías de calificación: T0-20, T21-40, T41-60, T61-80 y T81-100
Deterioros por longitud de puente	Numérica	Número de deterioros por kilómetro.
Deterioros por superficie de tablero	Numérica	Número de deterioros por área del tablero medida en kilómetros cuadrados.
Categoría de deterioro	Categórica	Etiquetas de categorías de deterioro propuestas: Deterioro Estético, Deterioro No estructural, Deterioro Estructural Menor, Deterioro Estructural Mayor.

En la Tabla 5.7 se muestran las 29 variables explicativas incluidas en el análisis. La variable objetivo fue la calificación de condición y fue expresada como variable numérica (de 1 a 100) y como variable categórica segmentada según la Tabla 2.1, donde se aprecian cinco categorías de calificación.

Tabla 5.7 Variables incluidas en la modelación

Variables incluidas en el modelo	Tipo de Variable
Clase de ruta	Categórica
TMDA	Numérica
TMDA de vehículos livianos	Numérica
TMDA de vehículos pesados	Numérica
Proporción de vehículos livianos	Numérica
Proporción de vehículos pesados	Numérica
Tipología general	Categórica
Tipología estructural	Categórica
Longitud total (m)	Numérica
Ancho total (m)	Numérica
Número de tramos	Numérica
Material de estribos	Categórica
Número de cepas	Numérica
Material de cepas	Categórica
Número de vigas	Numérica
Vigas	Categórica
Losa	Categórica
Espesor de losa (m)	Numérica
Carpeta	Categórica
Material menos resistente	Categórica
Número total de deterioros	Numérica
Deterioros por longitud	Numérica
Deterioros por superficie de tablero	Numérica
Deterioro	Categórica
Categoría de deterioro	Categórica
Causa	Categórica
Gravedad	Categórica
Extensión	Categórica

Unidad	Categorica
--------	------------

5.5 Depuración de bases de datos unificadas

La depuración de las bases de datos consiste en excluir las variables y los registros que no cuenten con la información necesaria para ser parte del análisis con herramientas de ML.

Primero se identificó la cantidad de registros con datos faltantes para cada variable en las bases de datos unificadas por región. Se excluyeron de la modelación las variables que presenten más de un 50% de vacíos, indistintamente de a qué región pertenezcan: año de construcción, ancho de calzada, ancho de pasillo, tipo de fundación de cepas, tipo de fundación de estribos y material de fundación de estribos. En la A5.1 y la A5.2 del Anexo 5 se muestran las variables de las bases de datos unificadas, la cantidad de registros que no presentan información en esa variable y el porcentaje de registros sin información respecto al total de registros en cada región.

Luego se tomó en cuenta las condiciones mínimas que deben tener los registros para crear las variables nuevas propuestas. Se excluyeron los registros que no cuenten con los datos de TMDA, los registros que no cuenten con los datos de calificación de condición, número de deterioros y/o tipo de deterioro y los registros que no cuenten con las dimensiones del puente.

En la Tabla 5.8 se muestra la cantidad de registros totales en las bases de datos unificadas por región, la cantidad de registros excluidos por cada criterio de depuración y la cantidad de registros utilizados en las bases de datos para el análisis por región, todas tanto en número como en porcentaje respecto al total de registros en cada base de datos.

Tabla 5.8 Cantidad de registros excluidos en depuración

Región	Descripción de registros cuantificados	Cantidad de registros	Porcentaje de registros totales en la región (%)
Zona centro	Registros totales en bases de datos unificadas	4041	100
	Registros excluidos por falta de inspección	1	0
	Registros excluidos por ausencia de dimensiones	184	4
	Registros excluidos por ausencia de TMDA	875	22
	Registros utilizados en base de datos de modelación	2981	74
Zona sur	Registros totales en bases de datos unificadas	2999	100
	Registros excluidos por falta de inspección	97	3
	Registros excluidos por ausencia de dimensiones	0	0
	Registros excluidos por ausencia de TMDA	1035	35
	Registros utilizados en base de datos de modelación	1867	62

La depuración de registros reduce las bases de datos de las regiones de la zona centro y la zona sur a un 74% y a un 62% de la cantidad de registros originales respectivamente. Por regla general los algoritmos de ML sugieren ser entrenados con 10 registros por cada variable explicativa, ambas bases de datos superan ese umbral sin problemas para el entrenamiento.

5.6 Conclusiones

La información de las bases de datos de tráfico fue incluida en las bases de datos de inventario e inspección de puentes, generando un total de 58 variables con 4041 registros en la región de la zona centro y 2999 en la región de la zona sur, incluyendo variables numéricas, categóricas nominales y categóricas ordinales. Debido a la diversidad de variables es necesario determinar su influencia en la calificación de condición de puentes antes de generar un modelo de comportamiento de puentes.

De la estadística descriptiva de las variables es esperable que la tipología estructural de puentes no sea una variable significativa ya que se superponen sus distribuciones de calificación de condición visualizadas en diagramas de caja. La región de la zona sur concentra menores calificaciones de

condición en elementos de madera en comparación con la región de la zona centro. Es esperable que la influencia de la materialidad de puentes sea mayor en la región de la zona sur que en la región de la zona centro. Además, el 65% de los deterioros son de extensión pequeña y gravedad mínima, por lo que se espera que estas variables no sean influyentes.

La estadística descriptiva permite también orientar la creación de nuevas variables para incluirlas en el análisis de influencia. Es esperable que la relación entre cantidad de deterioros con las dimensiones del puente sean una variable influyente ya que expresa la densidad espacial de los deterioros.

La depuración de las bases de datos unificadas concluyó en un total de 2981 y 1867 registros para las regiones de la zona centro y sur respectivamente. Tener miles de registros para el entrenamiento de las técnicas de ML satisface la regla general de presentar 10 registros por cada variable explicativa incluida en la modelación. De esta forma, las técnicas tienen reconocer un patrón en la influencia de variables a partir una muestra extensa.

6 ANÁLISIS CON HERRAMIENTAS DE MACHINE LEARNING

6.1 Introducción

En este capítulo se detalla cómo se selecciona las técnicas de ML usadas para medir la influencia relativa de las variables explicativas en la calificación de condición de puentes, el proceso de modelación de las técnicas y los resultados en forma de ranking de influencia de variables.

6.2 Criterios de selección de técnicas

Se seleccionaron técnicas de aprendizaje supervisado para clasificar o predecir la calificación de condición, estas en conjunto con la herramienta de Importancia de Atributos por Permutación, entregan una medición de influencia de las variables explicativas. Las técnicas utilizadas para el análisis deben cumplir con los siguientes criterios de selección como inclusión de variables numéricas y categóricas simultáneamente, falta de relación directa entre variables dependientes e independientes, parámetros de modelación de baja complejidad y disponibilidad en software accesible.

Las técnicas de regresión lineal y regresión logística no fueron utilizadas por perder interpretabilidad en problemas de falta de relación directa entre variables dependientes e independientes. Las técnicas de k-vecinos cercanos no fue utilizada por la falta de interpretabilidad de sus resultados y, en conjunto con Adaptive Boosting, por ser sensible a dato desbalanceados. Las técnicas de redes neuronales y las máquinas de soporte vectorial fueron descartadas por presentar parámetros de ajuste complejos.

6.3 Técnicas seleccionadas

Se seleccionaron tres técnicas: ReliefF (técnica basada en la asignación de un coeficiente de ponderación a las variables que diferencien a las clasificaciones entre sí), Random Forest (conjunto de árboles de decisión que votan un resultado entre sí) y Extreme Gradient Boosting (conjunto de árboles de decisión que se agregan ajustado los residuos o clasificaciones incorrectas anteriores). Las tres técnicas cumplen los criterios de selección y se encuentran presentes en un mismo software de programación. Se describió el funcionamiento de las técnicas y sus ventajas a continuación.

6.3.1 ReliefF

Es una técnica de asignación de coeficientes de influencia a variables explicativas en problemas de clasificación, basada en la técnica k-Vecinos Cercanos, no necesita la técnica de Importancia de

Atributos por Permutación. La técnica selecciona un registro fijo aleatorio y encuentra los “k” vecinos cercanos a este, luego compara el contenido de las variables explicativas entre los registros. Si los registros tienen distinta clasificación y el contenido de las variables explicativas difiere entonces aumenta el coeficiente de influencia de las variables explicativas. Este proceso se repite hasta que todos los registros hayan sido elegidos como registro fijo.

La técnica entrega como resultado coeficientes de influencia para cada variable explicativa sin necesidad de otra técnica complementaria. Compara múltiples variables al mismo tiempo ya sean numéricas o categóricas sin necesidad de separarlas. Además, no es sensible a datos vacíos mientras la mayoría de los registros no presente datos vacíos (Urbanowicz et al., 2018). Su único parámetro de ajuste es la cantidad de vecinos incluidos en la técnica de k-Vecinos Cercanos para determinar la cantidad de vecinos a comparar.

6.3.2 Random Forest

Técnica de clasificación o regresión que utiliza un conjunto de árboles de decisión generados desde distintas muestras del conjunto de registros de entrenamiento. Una vez generados, los distintos árboles de decisión clasifican o predicen una variable objetivo independientemente y luego votan entre sí para generar un resultado que prevenga el sobreajuste. Los árboles de decisión se caracterizan por incluir un gran número de variables explicativas jerarquizando su participación en la etapa de entrenamiento, permitiendo modelar patrones con falta de relación directa entre variables independientes y dependiente.

Los árboles de decisión permiten el uso de las variables categóricas y numéricas en el modelo, por lo que es posible comparar el efecto de las variables sin separarlas por tipo en distintas técnicas. Puede ser usado para tareas de clasificación o regresión (variable objetivo numérica o categórica). Presenta parámetros de ajuste fáciles de configurar como cantidad de árboles a generar, profundidad de niveles máximo en los árboles y cantidad de registros mínimos por hoja (nodo final del árbol). Es necesario utilizar la técnica en conjunto con Importancia de Atributos por Permutación para medir la influencia de las variables.

6.3.3 Extreme Gradient Boosting

También llamada XGBoost, es una técnica de clasificación o regresión supervisada, basada en un conjunto de árboles de decisión que ajusta de gradualmente los residuos o clasificaciones erróneas del árbol anterior. Incluye las ventajas de los árboles de decisión respecto a las variables explicativas numéricas o categóricas utilizadas jerárquicamente de forma simultánea.

Permite incluir registros con registros con datos faltantes en su entrenamiento (Sun et al., 2020). Presenta parámetros de ajuste similares a los de *Random Forest*, profundidad de niveles y cantidad de registros mínimos por hoja (nodo final del árbol), además de una tasa de aprendizaje, que asigna la proporción en que se van agregando los ajustes a los residuos o errores de clasificación. Es necesario utilizar la técnica en conjunto con Importancia de Atributos por Permutación para medir la influencia de las variables.

6.4 Aplicación de técnicas

Las técnicas seleccionadas se modelaron en Orange 3.34 (Demsar et al., 2013), software gratuito de programación basada en objetos que utiliza librerías del lenguaje de programación Python para entrenar modelos de ML. El software se basa en crear un flujo de trabajo de pequeñas aplicaciones incluidas en Orange (*Widgets*) con distintas con tareas específicas conectadas, entre ellas la edición de datos, el entrenamiento de técnicas de ML y la visualización de resultados de las técnicas.

En la Figura 6.1 se muestra el flujo de trabajo modelado con distintos bloques de tareas. El modelo incluye el proceso de selección de variables a modelar, la asignación de variables objetivo numérica y categórica, el entrenamiento de técnicas de ML y la medición de influencia de las variables explicativas.

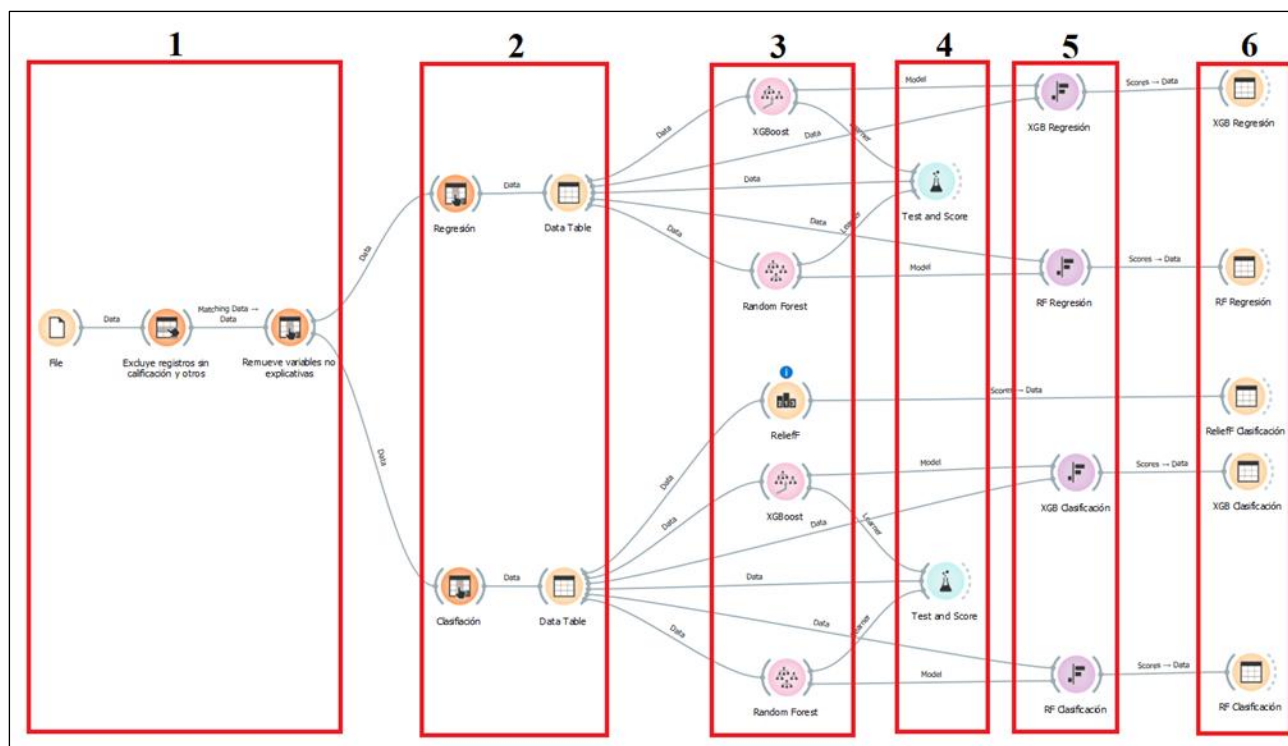


Figura 6.1 Implementación de técnicas en Orange

En el bloque 1 de la Figura 6.1 se muestran tres *widgets*, en el primero se ingresó la base de datos unificada, en el segundo se excluyeron los registros que no presenten información necesaria para la modelación como se explica en el Capítulo 5.5 y en el tercero se seleccionaron exclusivamente las variables explicativas seleccionadas en el Capítulo 5.4. En el bloque 2 con los *widgets* de la izquierda se asignó la variable objetivo de calificación de condición numérica en la parte superior y calificación de condición categórica en la parte inferior, ambos utilizaron la misma base de datos de entrenamiento, luego se envió la selección a un *widget* de tabla para los *widgets* de entrenamiento de técnicas.

En el bloque 3 de la Figura 6.1 se entrenan las técnicas Random Forest y Extreme Gradient Boosting como herramientas de regresión (calificación de condición numérica) en la parte superior. En la parte inferior se utiliza ReliefF y se entrenan las técnicas de Random Forest y Extreme Gradient Boosting (XGBoost) como herramientas de clasificación (calificación de condición categórica). En el bloque 4 se mide el rendimiento de las técnicas de ML, en este se utiliza la validación cruzada en cinco segmentos para prevenir el sobreajuste. En el bloque 5 se implementa la técnica de Importancia de Atributos por Permutación aplicada en Random Forest y XGBoost, generando un puntaje de influencia

para las variables explicativas expresado en figuras. En el bloque 6 se extrae los puntajes de influencia de las figuras del bloque 5 en forma de tablas.

En la Tabla 6.1 se muestran los resultados de rendimiento de los modelos. En problemas de clasificación, se midió el rendimiento de las técnicas según su Sensibilidad, que se define como la proporción de clasificaciones correctas sobre las instancias reales. En problemas de regresión se mide el rendimiento de las técnicas según el Coeficiente de determinación. Ambas mediciones de rendimiento son usadas como referencia para la Importancia de Atributos por Permutación.

Tabla 6.1 Rendimiento de técnicas en regresión y clasificación por región

Técnica de ML	Coeficiente de determinación (Regresión)		Sensibilidad (Clasificación)	
	Región Zona Centro	Región Zona Sur	Región Zona Centro	Región Zona Sur
Random Forest	0.992	0.964	0.995	0.964
Extreme Gradient Boosting	0.991	0.975	0.975	0.993

6.5 Influencia relativa de variables

Se mide la influencia de todas las variables explicativas utilizando la técnica ReliefF y la técnica de Importancia de Atributos por Permutación aplicada en Random Forest y XGBoost, tanto en tareas de regresión como clasificación. Se selecciona el conjunto de las variables más influyentes por técnica, tomando en cuenta solo las variables que presenten un puntaje de influencia de la misma escala de magnitud que la variable más influyente. En las Figura 6.2 a Figura 6.6 se muestran las variables más influyentes por técnica.

6.5.1 ReliefF

	ReliefF		ReliefF
C Clase de ruta	0.209	C Carpeta	0.331
C Vigas	0.203	C Clase de ruta	0.308
N Espesor de losa (m)	0.192	N Número total de deterioros	0.221
N Número de cepas	0.163	N Espesor de losa (m)	0.185
C Causa	0.128	N Proporción de vehículos livianos	0.183
C Unidad	0.122	N Proporción de vehículos pesados	0.183
N Número total de deterioros	0.120	N Número de cepas	0.177
C Deterioro	0.110	N Deterioros por longitud de puente	0.166
N Ancho total (m)	0.105	C Material de cepas	0.155
N Deterioros por longitud de puente	0.097	C Material de estribos	0.151
C Tipología estructural	0.090	C Tipología estructural	0.145
N Número de vigas	0.090	N TMDA de vehículos livianos	0.122
N TMDA de vehículos pesados	0.086	N Deterioros por superficie de tablero	0.122
N TMDA	0.083	N Número de vigas	0.116
C Carpeta	0.081	N Número de tramos	0.113
N TMDA de vehículos livianos	0.079	N TMDA	0.111
C Tipología general	0.076	C Losa	0.100
N Deterioros por superficie de tablero	0.075	N Longitud total (m)	0.091
C Material	0.072	N TMDA de vehículos pesados	0.089
N Número de tramos	0.070	C Material	0.079
N Proporción de vehículos pesados	0.062	C Vigas	0.076
N Proporción de vehículos livianos	0.062	N Ancho total (m)	0.062
N Longitud total (m)	0.057	C Tipología general	0.047
C Material de cepas	0.054	C Unidad	0.035
C Losa	0.029		

(a) Región zona centro

(b) Región zona sur

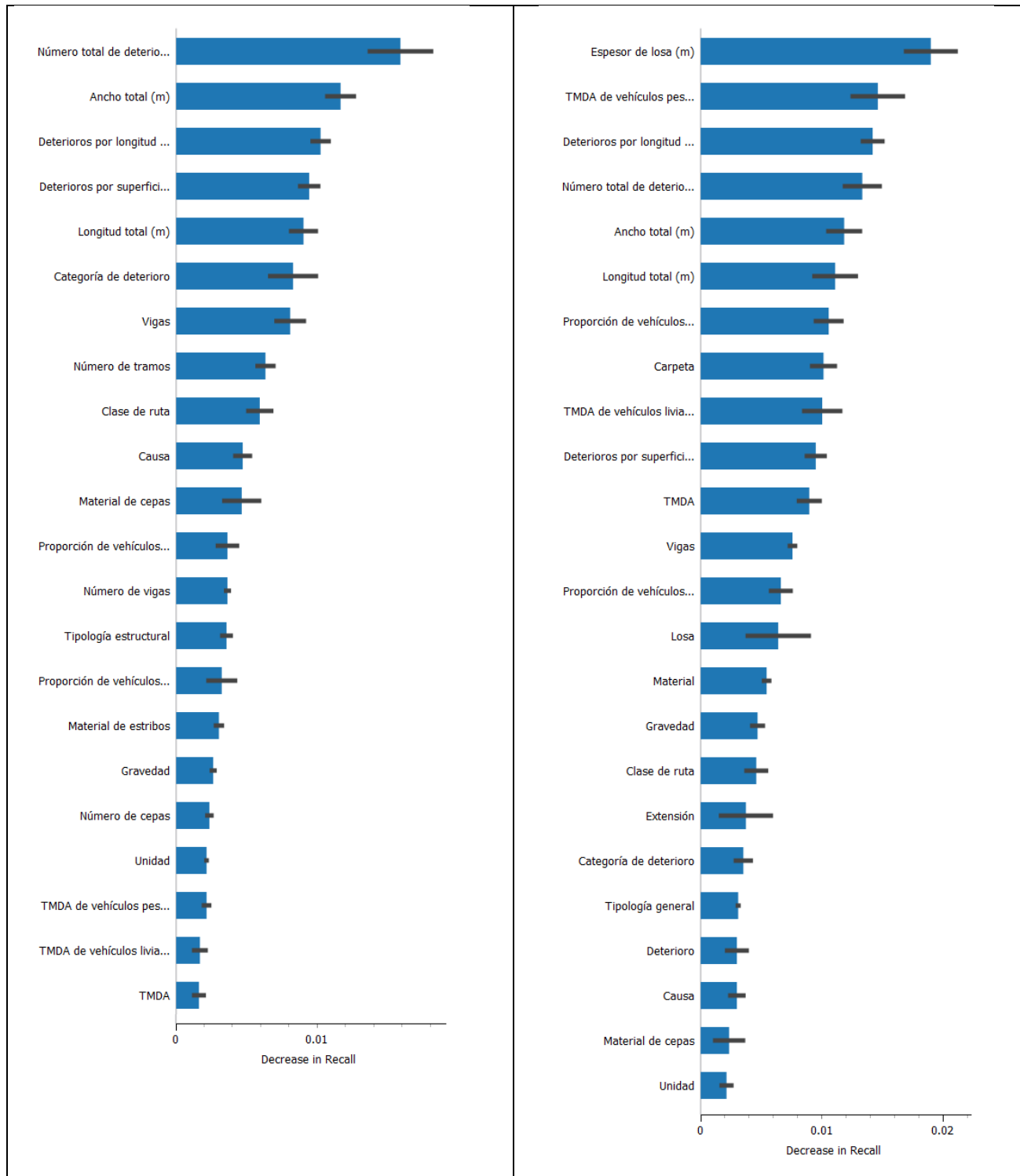
Figura 6.2 Ranking de variables influyentes obtenidas mediante ReliefF

En la Figura 6.2 (a) y (b) se muestra el nombre de las variables más influyentes en la primera columna, con una N roja si es una variable numérica y una C verde si es una variable categórica, en la segunda columna se ve el puntaje de influencia correspondiente a la técnica ReliefF.

En la región de la zona centro (Figura 6.2 (a)) las variables más influyentes fueron la clase de ruta, la materialidad de las vigas y el espesor de losa. En la región de la zona sur (Figura 6.2 (b)) las variables más influyentes fueron materialidad de la carpeta, la clase de ruta y el número total de deterioros. Para ambas regiones, la distribución de influencia de las variables decrece gradualmente entre las más influyentes a las menos influyentes, incluyendo 25 variables en la región de la zona centro y 24 variables en la región de la zona sur.

6.5.2 Random Forest y XGBoost en tareas de clasificación

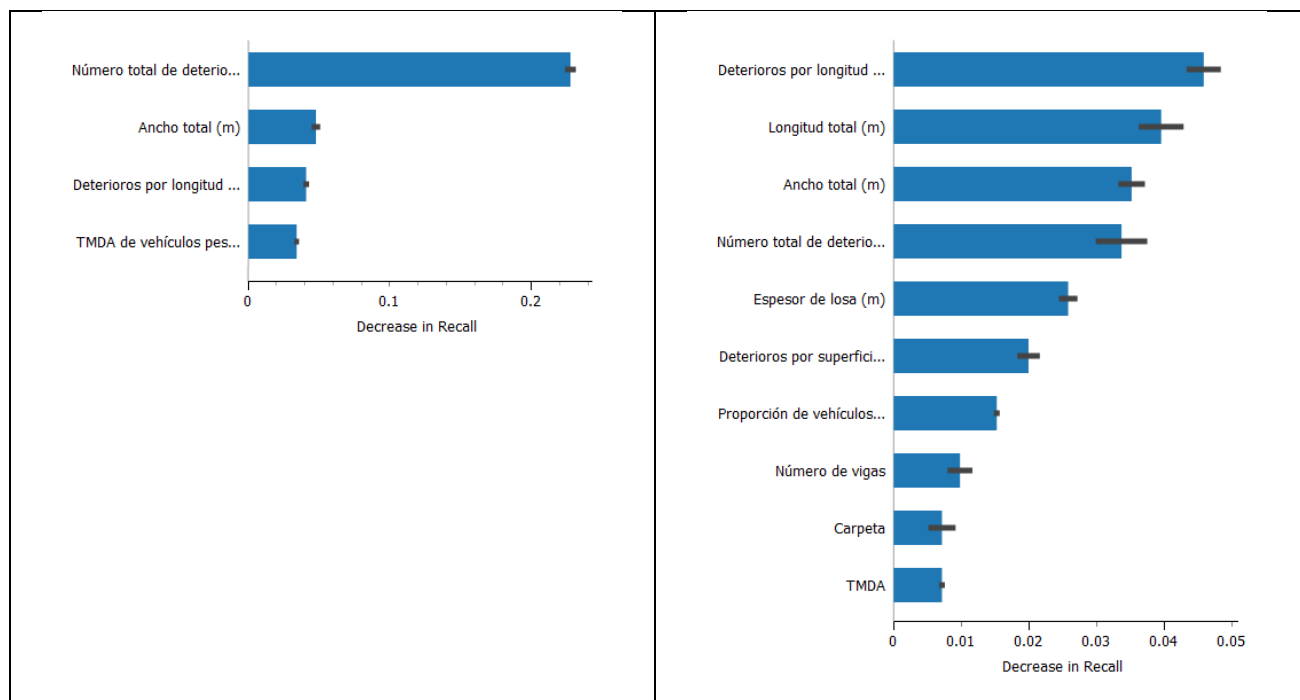
La Importancia de Atributos por Permutación (Brieman, 2001) mide la influencia de las variables en modelos ya entrenados. Intercambia aleatoriamente los valores de una variable entre registros y usa el modelo ya entrenado para clasificarlos o predecir su valor. Si al intercambiar los valores en la variable explicativa entre los registros se reduce el rendimiento de clasificación o predicción, entonces la variable explicativa realmente afecta a la variable objetivo. Se profundiza el funcionamiento de la técnica en la sección 3.3.



(a) Región zona centro (b) Región zona sur
Figura 6.3 Ranking de variables influyentes obtenidas mediante Random Forest (en tareas de clasificación)

En la Figura 6.3 (a) se muestra en el eje horizontal la reducción del rendimiento medido en *Recall* (o bien Sensibilidad) del modelo para cada variable, la escala de reducción de rendimiento llega hasta 0,02 aproximadamente, similar a la magnitud en la región de la zona sur. Las variables de mayor influencia son número total de deterioros, el ancho total del puente y la cantidad de deterioros por longitud de puente. La influencia de las variables decrece gradualmente entre desde la tercera variable más influyente en adelante, incluyendo 22 variables en el ranking.

En la Figura 6.3 (b) la escala de reducción de rendimiento medido en Sensibilidad llega hasta 0,02 aproximadamente, similar magnitud en la técnica de Random Forest para clasificación en la región de la zona centro. Las variables que al ser desordenadas (mostrando muestren la mayor reducción de rendimiento como espesor de losa, TMDA de vehículos pesados y cantidad de deterioros por longitud de puente son las más influyentes. La influencia de las variables decrece gradualmente entre las variables más influyentes a las menos influyentes, incluyendo 24 variables en el ranking.



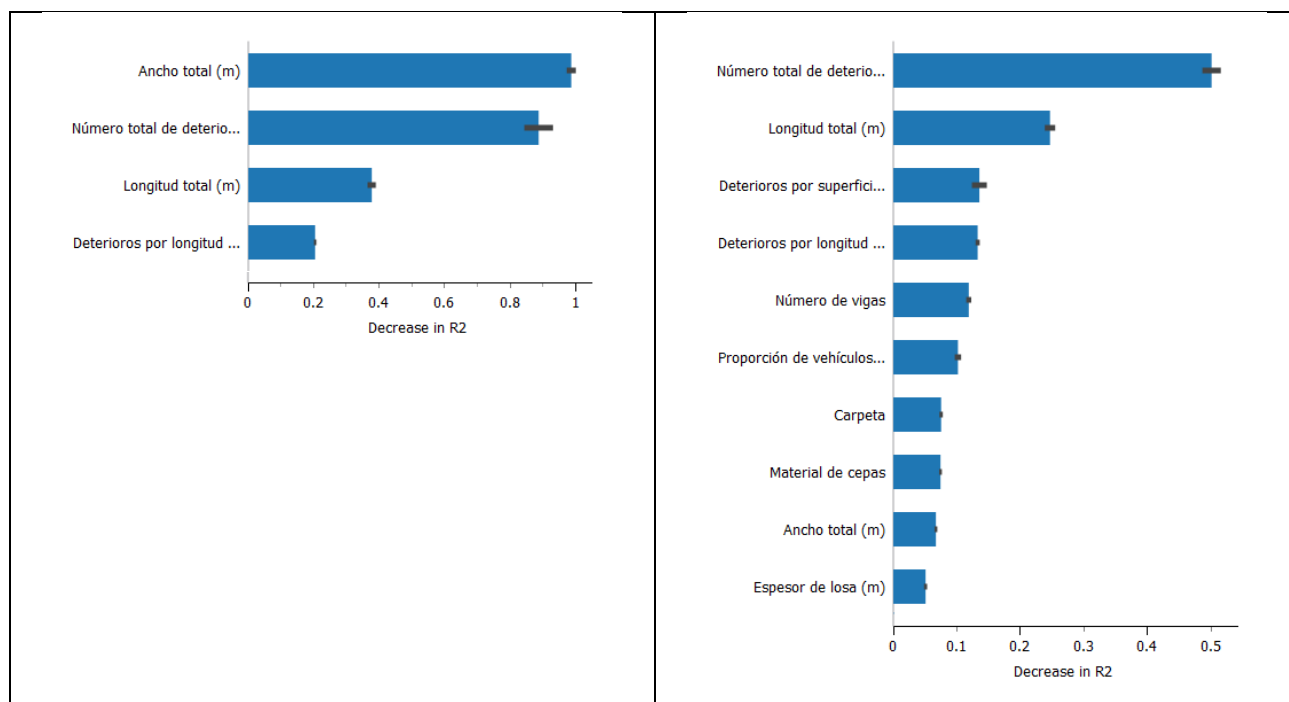
(a) Región zona centro **(b) Región zona sur**
Figura 6.4 Ranking de variables influyentes obtenidas mediante XGBoost (en tareas de clasificación)

En la Figura 6.4 (a) se muestra en el eje horizontal la reducción del rendimiento medido en *Recall* (o bien Sensibilidad) del modelo para cada variable. La escala de reducción de rendimiento llega hasta

0,25 aproximadamente, magnitud mayor que en la técnica de Random Forest para clasificación en ambas regiones. Las variables número total de deterioros, ancho total de puente, cantidad de deterioros por longitud de puente y TMDA de vehículos pesados fueron las más influyentes. La influencia de las variables decrece rápidamente entre la variable más influyentes a las menos influyentes, incluyendo solo cuatro variables en comparación con la mayoría de las variables incluidas en las técnicas ReliefF y Random Forest. El resto de las variables explicativas presenta una escala de puntaje de influencia menor por lo que no son incluidas en el ranking.

En la Figura 6.4 (b) se muestra en el eje horizontal la reducción del rendimiento medido en *Recall* (o bien Sensibilidad) del modelo para cada variable. La escala de reducción de rendimiento llega hasta 0.05 aproximadamente, magnitud menor que en la técnica de Extreme Gradient Boosting en la región de la zona centro, pero similar a la magnitud de Random Forest para clasificación en ambas regiones. Las variables más influyentes fueron la cantidad de deterioros por longitud de puente, la longitud total del puente y el ancho total del puente. La influencia de las variables decrece gradualmente desde las variables más influyentes a las menos influyentes, incluyendo diez variables en el ranking. Se incluyen más variables que en la región de la zona centro, pero todavía no incluye la mayoría de las variables, a diferencia de la técnica Random Forest y ReliefF.

6.5.3 Random Forest y XGBoost en tareas de regresión



(a) Región zona centro

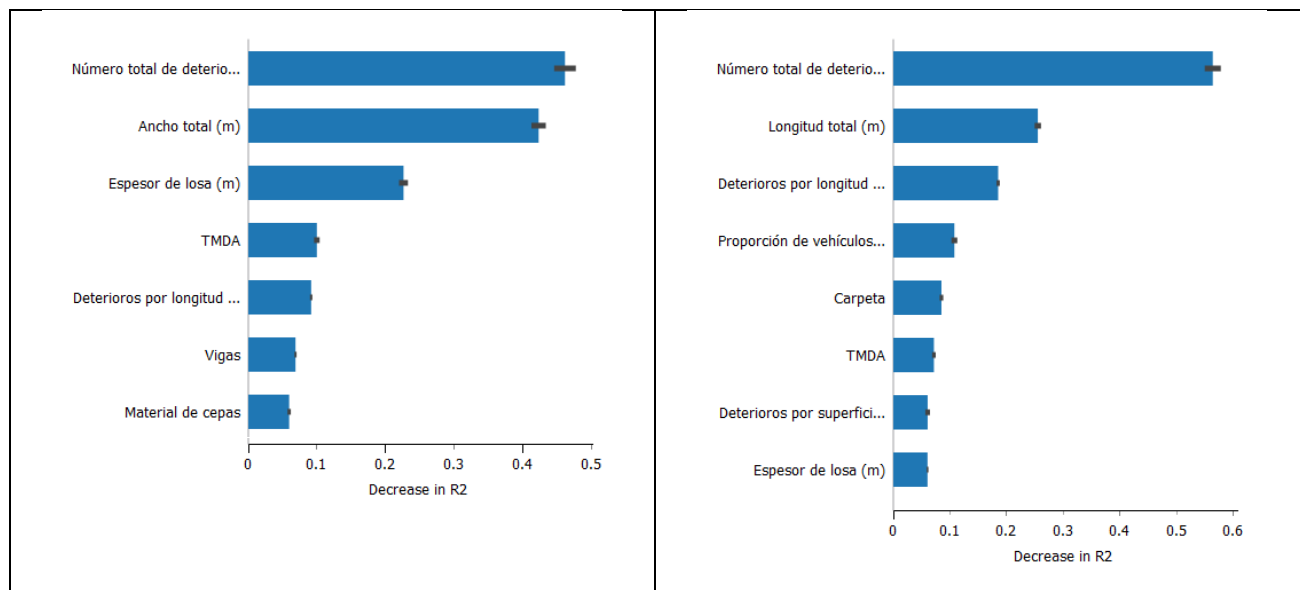
(b) Región zona sur

Figura 6.5 Ranking de variables influyentes obtenidas mediante Random Forest (en tareas de regresión)

En la Figura 6.5 (a) se muestra en el eje horizontal la reducción del rendimiento medido según el Coeficiente de determinación del modelo para cada variable. La escala de reducción de rendimiento llega hasta 1,00 aproximadamente, magnitud mayor que en la técnica de Random Forest en la región de la zona sur. Esta escala de magnitud de reducción de rendimiento es mucho mayor a la vista en todo el resto de las técnicas y podría ser indicador de sobreajuste, ya que asigna casi toda la capacidad de predicción del modelo a solo dos variables.

Las variables que al ser desordenadas muestran la mayor reducción de rendimiento como el ancho total de puente, el número de deterioros y la longitud total del puente fueron las más influyentes. La influencia de las variables decrece rápidamente desde la segunda variable más influyente, incluyendo solo cuatro variables en el ranking, situación similar a la de la técnica XGBoost en tareas de clasificación. El resto de las variables explicativas presenta una escala de puntaje de influencia menor por lo que no son incluidas en el ranking.

En la Figura 6.5 (b) se muestra en el eje horizontal la reducción del rendimiento medido según el Coeficiente de determinación del modelo para cada variable. La escala de reducción de rendimiento llega hasta 0,5 aproximadamente, magnitud menor que en la técnica la técnica de Random Forest en la región de la zona centro. Las variables que al ser desordenadas muestran la mayor reducción de rendimiento como el número de deterioros, la longitud total del puente y la cantidad de deterioros por superficie de tablero fueron las más influyentes. La influencia de las variables decrece gradualmente desde las variables más influyentes hacia las menos influyentes, incluyendo diez variables en el ranking.



(a) Región zona centro

(b) Región zona sur

Figura 6.6 Ranking de variables influyentes obtenidas mediante XGBoost (en tareas de regresión)

En la Figura 6.6 (a) se muestra en el eje horizontal la reducción del rendimiento medido según el Coeficiente de determinación del modelo para cada variable. La escala de reducción de rendimiento llega hasta 0,5 aproximadamente, similar a la magnitud de la técnica en la región de la zona sur. Las variables que al ser desordenadas muestran la mayor reducción de rendimiento como el número de deterioros, el ancho total del puente y el espesor de la losa fueron las más influyentes. La influencia de las variables rápidamente desde la tercera variable más influyente hacia las menos influyentes, incluyendo siete variables en el ranking.

En la Figura 6.6 (b) se muestra en el eje horizontal la reducción del rendimiento medido según el Coeficiente de determinación del modelo para cada variable. La escala de reducción de rendimiento llega hasta 0,6 aproximadamente, similar a la magnitud de la técnica la región de la zona centro. Las variables que al ser desordenadas muestran la mayor reducción de rendimiento como el número de deterioros, la longitud total del puente y el número de deterioros por longitud de puente fueron las más influyentes. La influencia de las variables rápidamente desde la tercera variable más influyente hacia las menos influyentes, incluyendo ocho variables en el ranking.

Existe una diferencia entre la magnitud de puntajes de influencia entre las distintas técnicas usadas y en ocasiones entre las mismas regiones. Fue esperable que existieran estas diferencias entre técnicas distintas, pero no en regiones distintas, como el caso de la Figura 6.5 (a) donde la magnitud de disminución de rendimiento indica sobreajuste del modelo. Fue necesario incluir las variables que presenten una magnitud de puntaje de medición similar por técnica, sino se corría el riesgo de incluir variables poco influyentes en el análisis.

6.5.4 Influencia promedio entre técnicas

Posteriormente se generaron dos conjuntos filtrados de variables influyentes, uno para tareas de clasificación y otro para tareas de regresión. Se incluyeron las variables que se hayan presentado como influyentes en una técnica o más. Este proceso resulta en conjuntos de 19 y 20 variables influyentes en tareas de clasificación para las regiones de la zona centro y sur respectivamente, y conjuntos de 9 y 11 variables influyentes en tareas de regresión para las regiones de la zona centro y sur respectivamente.

Se midió la influencia de las variables filtradas (detalle en el ANEXO 6.1 Ranking de influencia de variables), utilizando las técnicas de Importancia de Atributos por Permutación aplicadas en XGBoost y Random Forest y la técnica ReliefF. Las magnitudes de medición difieren para cada técnica, por lo que las mediciones de influencia fueron normalizadas entre la mínima y la máxima medición de influencia para cada técnica en forma porcentual. Entre las Figura 6.7 y Figura 6.10 se muestra en la primera columna el nombre de variables más influyentes en técnicas de clasificación o regresión por región. Los títulos de las columnas indican la técnica usada para medir la influencia de las variables. Los porcentajes muestran la medición de influencia normalizada por tipo de técnica.

Variable	ReliefF	XGBoost	Random Forest
Clase de ruta	100%	2%	25%
TMDA	32%	2%	6%
TMDA de vehículos livianos	28%	0%	21%
TMDA de vehículos pesados	38%	23%	15%
Proporción de vehículos livianos	32%	6%	17%
Proporción de vehículos pesados	32%	0%	15%
Tipología estructural	89%	0%	6%
Longitud total (m)	15%	5%	37%
Ancho total (m)	35%	25%	31%
Número de tramos	22%	4%	12%
Número de cepas	66%	0%	0%
Material de cepas	31%	0%	49%
Número de vigas	44%	1%	12%
Vigas	58%	8%	92%
Número total de deterioros	54%	100%	100%
Deterioros por longitud de puente	34%	22%	45%
Deterioros por superficie de tablero	30%	7%	36%
Causa	5%	2%	18%
Unidad	0%	2%	9%

Figura 6.7 Ranking de variables filtradas (clasificación) - Región zona centro

En la Figura 6.7 se muestra que los puntajes de influencia normalizados en la técnica ReliefF son más equiparados entre variables que las otras técnicas. En XGBoost la variable más influyente de número total de deterioros (100%) es cuatro veces mayor que la influencia de la segunda variable más influyente de ancho total de puente (25%), de la cuarta variable más influyente en adelante la influencia normalizada no es de más del 7%. En Random Forest los puntajes de influencia normalizados son más equiparados que en XGBoost, la variable más influyente es la cantidad total de deterioros (100%), le sigue la materialidad de las vigas (92%), el material de las cepas (49%) y la cantidad de deterioros por longitud del puente (45%), el resto de variables no superan una influencia normalizada mayor al 40%.

Variable	Relieff	XGBoost	Random Forest
Clase de ruta	100%	20%	9%
TMDA	72%	2%	32%
TMDA de vehículos livianos	76%	3%	26%
TMDA de vehículos pesados	67%	0%	25%
Proporción de vehículos livianos	83%	42%	30%
Proporción de vehículos pesados	83%	0%	32%
Tipología general	47%	17%	0%
Longitud total (m)	76%	96%	34%
Ancho total (m)	63%	97%	19%
Material de cepas	87%	0%	28%
Número de vigas	83%	6%	16%
Vigas	68%	0%	20%
Losa	68%	0%	75%
Espesor de losa (m)	88%	45%	75%
Carpeta	97%	12%	77%
Material	63%	0%	8%
Número total de deterioros	97%	100%	100%
Deterioros por longitud de puente	92%	60%	58%
Deterioros por superficie de tablero	80%	43%	30%
Unidad	0%	1%	14%

Figura 6.8 Ranking de variables filtradas (clasificación) - Región zona sur

En la Figura 6.8 los puntajes de influencia normalizados presentan valores más homogéneos entre sí por técnica, en comparación con la región de la zona centro. La variable de número total de deterioros y de longitud total del puente presentan mediciones de influencia normalizada altas en más de una técnica.

En XGBoost y Random Forest los puntajes de influencia normalizados se encuentran más equiparados en comparación a la región de la zona centro. XGBoost indica una influencia marcada en las dimensiones de los puentes, destacando el ancho total (97%), la longitud total (96%) y la cantidad de deterioros por longitud de puente (60%). Mientras que en Random Forest se destaca la materialidad de carpeta (77%), la materialidad y espesor de losa (75%).

Variable	XGBoost	Random Forest
TMDA	14%	15%
TMDA de vehículos pesados	0%	5%
Longitud total (m)	10%	35%
Ancho total (m)	92%	100%
Material de cepas	23%	0%
Vigas	13%	2%
Espesor de losa (m)	39%	4%
Número total de deterioros	100%	86%
Deterioros por longitud de puente	17%	23%

Figura 6.9 Ranking de variables filtradas (regresión) - Región zona centro

En la Figura 6.9 los puntajes normalizados de influencia se concentran en dos variables explicativas, el ancho total del puente (92% para XGBoost y 100% para Random Forest) y el número de deterioros presentes (100% para XGBoost y 86% y Random Forest). El resto de las variables no supera un puntaje de influencia normalizado del 40%.

Variable	XGBoost	Random Forest
TMDA	4%	16%
Proporción de vehículos livianos	11%	24%
Longitud total (m)	62%	22%
Ancho total (m)	2%	16%
Material de cepas	2%	0%
Número de vigas	8%	18%
Espesor de losa (m)	0%	0%
Carpeta	4%	7%
Número total de deterioros	100%	100%
Deterioros por longitud de puente	17%	56%
Deterioros por superficie de tablero	5%	26%

Figura 6.10 Ranking de variables filtradas (regresión) - Región zona sur

En la Figura 6.10, la distribución de influencia se concentra en las variables de número total de deterioros (100% para ambas técnicas), en la longitud del puente (62% para XGBoost) y en la cantidad de deterioros por longitud de puente (56% para Random Forest).

Por último, se muestra la influencia normalizada promedio de cada variable explicativa en la Figura 6.11, siendo la variable más influyente el número total de deterioros con una medición de influencia promedio normalizada entre técnicas de 94%, seguido por el ancho total con un 49% y la materialidad

de la losa con un 48%. Se excluyeron del ranking las variables que no estaban entre las variables más influyentes por cada técnica en las secciones anteriores. La influencia normalizada de las variables permite compararlas entre sí, independiente de los puntajes de influencia resultantes de cada técnica aplicada. El ranking de la Figura 6.11 muestra que la variable más influyente (número total de deterioros) presenta un puntaje de influencia normalizado de casi el doble que la segunda variable más influyente (el ancho total del puente). Desde la segunda posición del ranking la influencia decrece gradualmente hacia la variable menos influyente.

Nombre de variable	Influencia promedio
Número total de deterioros	94%
Ancho total (m)	49%
Losa	48%
Clase de ruta	43%
Deterioros por longitud de puente	40%
Longitud total (m)	38%
Carpeta	33%
Tipología estructural	31%
Espesor de losa (m)	30%
Deterioros por superficie de tablero	30%
Vigas	30%
Proporción de vehículos livianos	29%
Proporción de vehículos pesados	27%
TMDA de vehículos livianos	26%
Material	24%
Número de vigas	22%
Número de cepas	22%
Tipología general	21%
Material de cepas	19%
TMDA de vehículos pesados	19%
TMDA	18%
Número de tramos	13%
Causa	8%
Unidad	4%

Figura 6.11 Influencia promedio de variables principales

6.6 Conclusiones

Las técnicas de ML seleccionadas de ReliefF, Random Forest y Extreme Gradient Boosting permitieron analizar las características de las bases de datos presentes en la memoria, incluyendo distintas variables numéricas y categóricas en la medición de influencia, la falta de relación directa entre variables dependiente e independientes y con pocos parámetros de ajuste en software accesible.

Las mediciones de influencia relativa por cada técnica difieren en unidad de medición entre ReliefF respecto a Random Forest y Extreme Gradient Boosting, las últimas dos diferían a su vez en unidad de medición y en magnitud para tareas de clasificación o regresión. Seleccionar las variables más influyentes por técnica permitió filtrar las variables de influencia irrelevante del conjunto de modelación. Luego generar las mediciones de influencia en un conjunto filtrado de variables y normalizar sus resultados permitieron comparar las distintas técnicas entre sí, para promediarlas determinando la influencia de cada variable explicativa en la calificación de condición de puentes.

La técnica ReliefF genera puntajes de influencia de variables más graduales que la técnica de Importancia de Atributos por Permutación, pero indica mayor importancia en variables que no necesariamente se conectan con la condición de puentes como la clase de ruta. En contraste, la Importancia de Atributos por Permutación indica variables influyentes que sí coinciden con el criterio ingenieril como la cantidad de deterioros, pero concentra los puntajes de influencia en las primeras variables explicativas, dificultando la interpretación de la influencia de las variables con menor puntuación. Estas diferencias entre las técnicas se regulan una a la otra al momento de generar el promedio de los puntajes normalizados de influencia, unificando las ventajas de cada técnica y disminuyendo sus desventajas.

La variable más relevante fue el número de deterioros con una medición de influencia promedio normalizada entre técnicas de 94%. Las características de ancho total (49%) y materialidad de losa (48%), la clase de ruta (43%) y la cantidad de deterioros por longitud de puente (40%) fueron de las más relevantes.

Las regiones presentan similitudes en el ranking de influencia de variables explicativas, ambas mantienen en los primeros lugares la cantidad de deterioros y las dimensiones de puentes de largo y

ancho totales (en algunos casos también espesor de losa). Además, en las variables que indican materialidad se repite la materialidad de vigas y cepas, pero no de estribos o de losa.

7 CONCLUSIONES Y RECOMENDACIONES

El propósito de esta memoria fue medir la influencia de distintas variables explicativas en la calificación de condición de puentes en bases de datos de Chile usando técnicas de Machine Learning (ML). Se seleccionaron las técnicas ReliefF e Importancia de Atributos por Permutación aplicada en Random Forest y Extreme Gradient Boosting para medir la influencia de variables en la calificación de condición, utilizando bases de datos unificadas de tráfico y catastro e inspección de puentes en dos regiones de Chile. Finalmente se generó un ranking de influencia promedio para una mejor interpretación y así orientar futuras calibraciones de modelos de comportamiento de puentes. El trabajo fue desarrollado con las bases de datos disponibles en dos de las 16 regiones del país, por lo que sus resultados se limitan a ese contexto.

La calidad de la información que se utilizó para el análisis presenta diversas variables explicativas que son importantes para la confección de un modelo de comportamiento de puentes como las dimensiones del puente, la materialidad de sus elementos y el tráfico asociado a este. Pero no se incluyen otras variables que da cuenta del contexto completo del puente, por ejemplo, la edad del puente presenta demasiados campos vacíos, no se incluye registros de altura del puente en distintos puntos, las condiciones del terreno donde se encuentra emplazado, un historial de intervenciones de mantención o un registro de eventos adversos (ya sean naturales o antrópicos). Estas variables pueden ser influyentes en la modelación del comportamiento de puentes.

Entre las variables de mayor influencia en la calificación de condición de puentes se encuentra el número de deterioros, el ancho total de puentes, la materialidad de la losa, la clase de ruta, la cantidad de deterioros por longitud de puente, y la materialidad de carpeta. Se presenta una concentración en el tipo de variables más influyentes liderado por la cantidad de deterioros y las dimensiones del puente, seguido por la materialidad de distintos elementos.

No se encuentran entre las más influyentes el tipo de deterioros, ni la extensión o la gravedad de estos incluso si estas son las variables de entrada para el cálculo de la calificación de condición de puentes. Las dimensiones de puentes, la materialidad de sus elementos y el tráfico no se encuentran en el cálculo de la calificación de condición a pesar de que sí son influyentes.

El contar con la influencia relativa de las variables explicativas permite sugerir no solo la inclusión de estas en modelos de comportamiento de puentes sino orientar su ponderación en el proceso de calibración. Permitiendo reducir el sesgo de elección de variables según experiencia ingenieril, previniendo la inclusión de variables que no aporten realmente a la predicción de comportamiento y dificulten la interpretabilidad del modelo.

La inclusión de variables nuevas derivadas de las presentes en las bases de datos originales permitió caracterizar la relación entre variables explicativas y cómo incluirlas en modelos de comportamiento. Un ejemplo de esto fue la variable influyente de cantidad de deterioros por longitud de puente que muestra la densidad espacial de los deterioros, que resulta ser más influyente que la materialidad de las vigas.

Las técnicas seleccionadas ReliefF e Importancia de Atributos por Permutación (aplicada en XGBoost y Random Forest) permitieron medir la influencia relativa de las variables explicativas en la condición de puentes, aplicando ventajas de estos como la inclusión de variables numéricas y categóricas simultáneamente, o el trabajo con variables sin información en registros, adecuándose a las características de las bases de datos de tráfico, inventario e inspección de puentes presentes en la memoria.

Se recomienda para trabajos futuros incluir otras variables como la edad de los puentes o un registro de intervenciones de mantenimiento y eventos adversos. Estas podrían ser variables influyentes en el comportamiento de los puentes ya que se relacionan con el deterioro de los puentes a través del tiempo. Además, es posible ampliar el trabajo a distintas regiones de Chile para analizar si existe un patrón de influencia de variables ligada al contexto geográfico de los puentes. Incluir estos análisis ayudaría a determinar la influencia de distintos fenómenos en el comportamiento de puentes que se escapan del alcance de la memoria.

REFERENCIAS

Breiman, L. (2001) Random Forests. *Machine Learning*, 45(1), 5-32

El Mountassir, M. E., Yaacoubi, S., Ragot, J., Maquin, D. (2016). Feature selection techniques for identifying the most relevant damage indices in SHM using guided waves. *e-Journal of Nondestructive Testing*, 21(8): 1-8 <https://www.ndt.net/?id=20151>

Malekloo A., Ozer E., AlHamaydeh M., Girolami M. (2022) Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, 21(4): 1906-1955

De Solminihaç, H. (2018). Gestión de infraestructura vial (T. Echaveguren Navarro & M. A. Chamorro Giné (Eds.); Tercera edición). Ediciones Universidad Católica de Chile.

Demsar J., Curk T., Erjavec A., Gorup C., Hocevar T., Milutinovic M., Mozina M., Polajnar M., Toplak M., Staric A., Stajdohar M., Umek L., Zagar L., Zbontar J., Zitnik M., Zupan B. (2013) Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14(71): 2349–2353.

Dirección de Vialidad Ministerio de Obras Públicas. (2022). *Plan de puentes 2020-2030* (1.^a ed.). Ministerio de Obras Públicas de Chile.

Omar I., Khan M., Starr A. (2022) Compatibility and challenges in machine learning approach for structural crack assessment. *Structural Health Monitoring*, 21(5): 2481-2502

Sun, L., Shang, Z., Xia, Y., Bhowmick, S., & Nagarajaiah, S. (2020). Review of Bridge Structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Journal of Structural Engineering*, 146(5).

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal Of Biomedical Informatics*, 85: 189-203.

Zhu, J., & Wang, Y. (2021). Feature selection and deep learning for deterioration prediction of the bridges. *Journal of Performance of Constructed Facilities*, 35(6).

ANEXOS

ANEXO 1.1 Contribución a los Objetivos de Desarrollo Sostenible

Los Objetivos de Desarrollo Sostenible (ODS) constituyen un llamamiento universal a la acción para poner fin a la pobreza, proteger el planeta y mejorar las vidas y las perspectivas de las personas en todo el mundo. Seleccione a cuál o cuáles de los 17 ODS contribuye su trabajo de Memoria de Título:

- ODS-1 : Fin de la pobreza.
- ODS-2 : Hambre cero.
- ODS-3 : Salud y bienestar.
- ODS-4 : Educación de calidad.
- ODS-5 : Igualdad de género.
- ODS-6 : Agua limpia y saneamiento.
- ODS-7 : Energía asequible y no contaminante.
- ODS-8 : Trabajo decente y crecimiento económico.
- ODS-9 : Industria, innovación e infraestructura.
- ODS-10 : Reducción de las desigualdades.
- ODS-11 : Ciudades y comunidades sostenibles.
- ODS-12 : Producción y consumo responsables.
- ODS-13 : Acción por el clima.
- ODS-14 : Vida Submarina.
- ODS-15 : Vida de ecosistemas terrestres.
- ODS-16 : Paz, justicia e instituciones sólidas.
- ODS-17 : Alianzas para lograr los objetivos.

Vinculación

La memoria de título tiene por objetivo diagnosticar la influencia de distintas variables que determinan la condición de un puente, sugiriendo que variables que tengan mayor influencia sean incluidas en el cálculo de calificación de condición de puentes. El incluir variables de mayor influencia, siendo estas fáciles de medir y con un mínimo sesgo de parte del observador, optimiza el proceso de calificación y orienta futuros procesos de evaluación y estrategias de mantención de la infraestructura vial de puentes.

ANEXO 2.1 Contenido de variables de tráfico vial, inventario e inspección de puentes

Tabla A.2.1.1 Contenido de variables de Tráfico

Tráfico	Contenido	Descripción
1. CODIGO	Código alfanumérico	Código abreviado de la ruta del puente
2. KI	Numérico	Kilómetro de inicio del tramo de la ruta
3. KF	Numérico	Kilómetro de fin del tramo de la ruta
4. TRAMO	Numérico	Número de tramo de la ruta
5. TMDA	Numérico	Tráfico Medio Diario Anual del tramo de la ruta
6. P-A	Numérico	% de TMDA correspondiente a Autos en el tramo de la ruta
7. P-C	Numérico	% de TMDA correspondiente a Camionetas en el tramo de la ruta
8. P-C2E	Numérico	% de TMDA correspondiente a Camiones de dos ejes en el tramo de la ruta
9. P-SR	Numérico	% de TMDA correspondiente a Semiremolques en el tramo de la ruta
10. P-R	Numérico	% de TMDA correspondiente a Remolques en el tramo de la ruta
11. P-C+2E	Numérico	% de TMDA correspondiente a Camiones de más de dos ejes en el tramo de la ruta
12. P-B	Numérico	% de TMDA correspondiente a Buses en el tramo de la ruta
13. Veloc_tramo	Numérico	Velocidad del tramo de la ruta

Tabla A.2.1.2 Contenido de variables de inventario de puentes

Nombre de Variable	Contenido	Descripción
1. Nombre Puente	Texto	Nombre del puente en texto
2. Código Puente	Código alfa-numérico	Código de la ruta del puente

3. Código I3MOP	Código alfa-numérico	Código del puente
4. Región	Texto	Nombre de la Región donde se encuentra el puente
5. Provincia	Texto	Nombre de la Provincia donde se encuentra el puente
6. Comuna	Texto	Nombre de la Comuna donde se encuentra el puente
7. Tuición	<ul style="list-style-type: none"> • Dirección de Vialidad • DDQ-APSA • DDQ Ingenieros Consultores 	Tuición administrativa del puente
8. Código Ruta	Código alfa-numérico	Código abreviado de la ruta del puente
9. Clase Ruta	<ul style="list-style-type: none"> • Caminos nacionales • Caminos regionales principales • Caminos regionales comunales • Caminos regionales provinciales • Caminos regionales de acceso 	Clase de camino
10. KM	Numérico	Kilómetro de ruta donde se ubica el puente
11. Tipología General	<ul style="list-style-type: none"> • Viaducto • Puente • No observable • Paso Inferior • Paso Superior • Modular 	Tipología del puente
12. Tipología Estructural	<ul style="list-style-type: none"> • Convencional vigas • Convencional cajón • Convencional losa • Convencional celosía • Colgante • Arco • Otros • No observable 	Tipología estructural del puente
13. Longitud total (m)	Numérico	

14. Ancho total (m)	Numérico	
15. Número de tramos	Numérico	
16. Año de construcción	Numérico	Formato de número, no de fecha
17. Número de Cepas	Numérico	
18. Ancho de Calzada (m)	Numérico	
19. Ancho de Pasillo (m)	Numérico	
20. Espesor de Losa (m)	Numérico	
21. Número de Vigas	Numérico	
22. Material cepas	<ul style="list-style-type: none"> • Hormigón • Hormigón armado • Hormigón en masa • Gaviones • Acero • Madera • No existe 	Material de cepas
23. Fundación cepas	<ul style="list-style-type: none"> • Profunda: cajón hincado • Superficial: zapata • Otra tipología • No observable • No existe 	Tipología de fundación de cepa
24. Material estribos	<ul style="list-style-type: none"> • Hormigón • Hormigón armado • Acero • Gaviones • Madera • Tierra reforzada • Otros materiales • No observable 	Material de estribos

	<ul style="list-style-type: none"> • No existe 	
25. Fundación estribos	<ul style="list-style-type: none"> • Superficial: zapata • Superficial: no determinada • Superficial: otra tipología • Profunda: cajón hincado • Profunda: encepado + pilotes • No observable 	Tipología de fundación de estribos
26. Material fundación estribos	<ul style="list-style-type: none"> • Hormigón armado • No observable 	Material de fundación de estribos
27. Carpeta	<ul style="list-style-type: none"> • Acero • Asfáltico • Hormigón • Hormigón armado • Hormigón en masa • Hormigón no determinado • Madera • Relleno granular • Tierra reforzada • Otros materiales • No aplica • No existe • No observable 	Material de carpeta
28. Vigas	<ul style="list-style-type: none"> • Hormigón postensado • Hormigón armado • Acero • Aluminio • Madera 	Material de vigas
29. Losa	<ul style="list-style-type: none"> • Hormigón • Hormigón armado • Hormigón no determinado • Acero • Madera • No Existe 	Material de losa

Tabla A.2.1.3 Contenido de variables de inspección de puentes

Inspección	Contenido	Descripción
1. Fecha inspección principal	Fecha	Fecha en formato de fecha de Excel
2. Calificación automática	Numérico	Calificación automática del puente
3. Calificación final	Numérico	Calificación final del puente (distinta de la automática dado que puede haber intervenciones de los inspectores)
4. Número total de deterioros	Numérico	
5. Índice Estructural	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
6. Índice Seguridad Vial	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
7. Índice Contención Sistema de Contención	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
8. Índice Contención Barandilla	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
9. Índice Durabilidad	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
10. Índice Daños Menores	Numérico	Índice interno del sistema I3MOP de cálculo desconocido
11. Deterioro	<ul style="list-style-type: none"> • Alteración superficial • Armadura vista • Baches, Ahuellamiento • Cárcava • Circulación o presencia de agua • Corrosión 	35 Tipos de deterioro específicos no adaptables

	<ul style="list-style-type: none">• Deformación• Depositación• Despegue• Desplazamiento• Desprendimientos de hormigón con armadura vista• Desprendimientos de hormigón sin armadura vista• Eflorescencias• Falta de alineación• Fisuras• Fisuras en mapa o en retícula• Grafiti• Grietas (>5 mm)• Hongos• Humedades, filtraciones• Hundimiento• Lajación• Nido de piedras• Obstrucción• Pátina, mancha de óxido• Pérdida de finos• Pérdida de pieza• Pérdida de tornillos, roblones, anclajes• Pérdida de tratamiento protector• Piezas sueltas• Pudrición / ataque de insectos• Restos metálicos de ejecución (clavos, barras)• Rotura• Socavación local• Vegetación	
--	--	--

12. Causa	<ul style="list-style-type: none">• Abrasión por tránsito• Acción climática• Acción de la fauna• Acción sísmica• Ataque químico• Ausencia o deficiente dispositivo de junta• Ausencia o deficiente losa de transición• Cabeceo del muro lateral y existencia de un punto fijo• Calidad deficiente de la pieza/elemento• Capilaridad• Causa desconocida• Ciclos hielo-deshielo• Corrosión• Corrosión de las armaduras• Deficiente compactación del relleno• Deficiente ejecución• Deformación natural del propio aparato de apoyo• Desgaste o abrasión por tráfico• Diseño deficiente• Envejecimiento• Escasez de recubrimiento• Escorrentía superficial• Esfuerzos• Esfuerzos (compresión)• Esfuerzos (cortante)	41 Tipos de causas de deterioros específicas no adaptables
-----------	--	--

	<ul style="list-style-type: none"> • Esfuerzos (flexión) • Esfuerzos (torsión) • Esfuerzos (tracción) • falta de barbacana • Falta de limpieza de encofrado • Falta de protección • Golpe o impacto • Impermeabilización defectuosa • Mal funcionamiento del sistema de drenaje • Perdida de tratamiento protector • Retracción • Sedimentación orgánica • Sedimentación por escorrentía superficial • Subestimación de empujes • Tensión excesiva en zonas de anclajes • Vandalismo 	
13. Gravedad	<ul style="list-style-type: none"> • Alta • Media • Mínima • Muy alta 	4 Categorías de gravedad de deterioro
14. Extensión	<ul style="list-style-type: none"> • Alta • Media • Mínima • Muy alta 	4 Categorías de extensión de deterioro
15. Medición	Numérico	Medición de deterioro asociado a una unidad de medida
16. Unidad medición	<ul style="list-style-type: none"> • cm • cm² • ml 	Unidad de medida de la medición del deterioro

	<ul style="list-style-type: none"> • m2 • m3 • Ud 	
17. Calificación automática del deterioro	Numérico	Calificación específica del deterioro en cuanto a su tipo, causa, extensión, gravedad y unidad donde afecta
18. Unidad	<ul style="list-style-type: none"> • Anillos • Aparato de apoyo • Arriostramiento horizontal • Arriostramiento vertical • Atiesador • Bajada de agua • Barbacanas • Barras sísmicas • Cable • Calzada Tramo I • Calzada Tramo II • Cama de nivelación • Cierre perimetral • Conexión - Sobre Tablero • Contraventación • Cordón inferior • Cordón superior • Diagonal • Dintel • Fundación • Horizontal • Longuerina • Losa • Losa de transición • Losa Superior • Macizo anclaje • Montante • Muro de contención de tierras • Muro espaldar • Muro frontal portante • Otras tuberías • Otros I • Pasillo Tramo I 	Unidad donde está ubicado el deterioro

	<ul style="list-style-type: none"> • Pasillos • Pavimento • Péndolas • Pilas - Sobre Tablero • Poste con luminaria • SC Vehículos • Solera • Tajamar • Terraplenes • Topes sísmicos • Transición al pavimento • Travesaño • Travesaño - Sobre Tablero • Travesaño central • Tuberías abastecim. • Tubos de desagüe • Vertical • Verticalidad (longitudinal) • Viga cabezal • Vigas • Voladizo 	
19. Código	Alfanumérico (Ej: E2 1, se refiere a: estribo 2 / primer elemento)	Código asociado a la unidad donde está asociado el deterioro

ANEXO 2.2 Cálculo de calificación de condición de puentes

En la ecuación 2.1 y la ecuación 2.2 (MOP, 2022) se aprecia el cálculo de la calificación de deterioro, donde se pondera su extensión y su gravedad (A2.2.1), su tipo (A2.2.2), su causa (A2.2.3) y la unidad a la que corresponde el deterioro (A2.2.4). En las tablas mencionadas se presentan extractos de los ponderadores influyentes en la calificación de condición de deterioro de puentes, aunque los desarrolladores no detallan como se obtuvieron los valores de ponderación.

$$Cd = Co \cdot Pd \cdot Pc \cdot Pu \quad (A2.1)$$

$$C_o = \frac{100 - (V_{Extensión} + V_{Gravedad}) \cdot 4}{100} \quad (A2.2)$$

Donde:

- Cd se refiere a la calificación de cada deterioro.
- V_{Gravedad} se define para cada deterioro y se le asigna un coeficiente de acuerdo con la A2.2.1
- V_{Extensión} se define para cada deterioro y se le asigna un coeficiente de acuerdo con la A2.2.1
- Pd se refiere al coeficiente de cada deterioro de acuerdo con la A2.2.2
- Pc se refiere al coeficiente de la causa de deterioro de acuerdo con la A2.2.3
- Pu se refiere al coeficiente de la unidad de un elemento afectado por el deterioro de acuerdo con la A2.2.4

Tabla A2.2.1 Ponderación de Extensión y Gravedad en Calificación de Deterioro I3MOP

Extensión	V_{Extensión}	Gravedad	V_{Gravedad}
Pequeña (<25%)	0,15	Mínima	0,10
Media (25% < Ext < 50%)	0,50	Media	3,00
Elevada (50% < Ext < 75%)	2,00	Alta	8,00
Muy elevada (>75%)	3,50	Muy alta	13,00

Fuente: Dirección de Vialidad, Ministerio de Obras Públicas [MOP].

Tabla A2.2.2 Ponderación de Tipo de Deterioro en Calificación de Deterioro I3MOP

Deterioro	Pd	Deterioro	Pd
Alteración superficial	97	Grietas (>5 mm)	87
Anclajes sueltos	93	Humedades, filtraciones	99

Armadura vista	96	Hundimiento	92
Depositación	96	Iluminación	99
Baches, ahuellamiento	97	Lajación	92
Cárcavas	96	Pérdida de finos	93
Circulación 0 presencia de agua	99	Madrigueras	98
Nido de piedras	99	Obstrucción	94
Corrosión	92	Pátina, mancha de óxido	99
Deformación	89	Pérdida de pieza	91
Socavación local	80	Pérdida de tornillos, roblones, anclajes	93
Desprendimientos de hormigón	92	Pérdida de tratamiento protector	98
Desprendimientos de hormigón sin armadura vista	99	Piezas sueltas	92
Despegue	95	Grafiti	100
Desplazamiento	93	Restos metálicos de ejecución (clavos, barras)	99
Eflorescencias	98	Rotura	85
Falta de alineación	99	Vegetación	99
Fisuras	88	Pudrición / ataque de insectos	93
Fisuras en mapa	93	Hongos	94

Fuente: Dirección de Vialidad, Ministerio de Obras Públicas [MOP]

**Tabla A2.2.3 Ponderación de Causa de Deterioro en Calificación de Deterioro
I3MOP**

Causa	Pc	Causa	Pc
Abrasión por tránsito	0,98	Exceso de compresión	0,95
Acción climática	0,98	Falta de compresión	0,95
Acción de la fauna	0,98	Falta de cuña en el neopreno	0,98
Asiento diferencial	0,85	Falta de barbacana	0,97

Ataque químico	0,90	Falta de limpieza de encofrado	1,00
Ausencia o deficiente dispositivo de junta	0,97	Falta de protección	0,98
Ausencia o deficiente losa de transición	0,97	Falta de protección de anclajes	1,00
Cabeceo del muro lateral y existencia de un punto fijo	0,95	Falta de protección frente a avenidas	0,91
Cama de nivelación en mal estado o mal ejecutada	0,94	Calidad deficiente de la pieza/elemento	0,96
Capilaridad	0,99	Giro	0,88
Causa desconocida	0,90	Giro del estribo	0,85
Ciclos hielo-deshielo	0,97	Giro del muro lateral	0,90
Corrosión	0,93	Golpe o impacto	0,95
Corrosión de las armaduras	0,92	Impermeabilización defectuosa	0,98
Deficiente compactación del relleno	0,93	Infradimensionamiento de apoyos	0,95
Deficiente ejecución	0,97	Infradimensionamiento del elemento	0,90
Deformación natural del propio aparato de apoyo	0,99	Mal funcionamiento del sistema de drenaje	0,97

Fuente: Dirección de Vialidad, Ministerio de Obras Públicas [MOP]

Tabla A2.2.4 Ponderación de Unidad en Calificación de Deterioro I3MOP

Subcomponente	Unidad	Pu
Superestructura (bóveda)	Bóveda	0,91
	Tímpano	0,94
	Bóveda (tímpano)	0,94
	Boquilla (tímpano)	0,96
	Montante (tímpano)	0,96
	Boquilla	0,96
	Voladizo	0,96
	Imposta	0,98
	Puntales	0,94

Revestimiento	0,98
Cordón superior	0,91
Cordón inferior	0,91
Diagonal	0,91
Montante	0,91
Voladizo	0,94
Losa	0,94

Fuente: Dirección de Vialidad, Ministerio de Obras Públicas [MOP]

Es en este procedimiento de cálculo donde se identifican los principales ponderadores dentro del algoritmo, relacionados a variables incluidas en las bases de datos presentes como la extensión y gravedad de un deterioro o el elemento al que corresponde el deterioro. Sin embargo, este procedimiento no abarca todas las variables disponibles, por ejemplo, no considera variables como la materialidad de cada componente, el tráfico presente, el tipo de camino al que pertenece o las dimensiones del puente.

ANEXO 5.1 Estadística descriptiva de variables

Tabla A5.1 Registros vacíos de variables originales – Región Zona Centro

Nombre de variable	Cantidad de registros sin datos	Porcentaje de registros sin datos respecto a total de registros en la región (%)
NOMBRE_PUENTE	0	0
CÓDIGO_PUENTE	608	15
CÓDIGO I3MOP	1	0
REGIÓN	0	0
PROVINCIA	26	1
COMUNA	26	1
TUICIÓN	318	8
CODIGO_RUTA	26	1
CLASE_RUTA	79	2
KM	7	0
Tramo	0	0

TMDA	943	23
P-A	943	23
P-C	943	23
P-C2E	943	23
P-SR	943	23
P-R	943	23
P-C+2E	943	23
P-B	943	23
Veloc_tramo	379	9
Tipología General	60	1
Tipología Estructural	106	3
Año de construcción	3479	86
Longitud total (m)	185	5
Ancho total (m)	185	5
Ancho de Calzada (m)	2198	54
Ancho de Pasillo (m)	4015	99
Número de tramos	61	2
Número de Cepas	1338	33
Material cepas	1368	34
Fundación cepas	3285	81
Material estribos	154	4
Fundación estribos	3243	80
Material fundación estribos	2274	56
Número de Vigas	570	14
Vigas	951	24
Losa	488	12
Espesor de Losa (m)	1375	34
Carpeta	130	3
Fecha última inspección principal	0	0
Calificación automática	217	5
Calificación final	0	0
Número total de deterioros	0	0
Índice Estructural	671	17
Índice Seguridad Vial	671	17
Índice Contención Sistema de Contención	671	17
Índice Contención Barandilla	671	17
Índice Durabilidad	671	17
Índice Daños Menores	671	17
Deterioro	1	0
Causa	255	6
Gravedad	1	0
Extensión	1	0
Medición	589	15

Unidad medición	594	15
Calificación automática de deterioro	37	1
Unidad	35	1
Código	856	21

Tabla A5.2 Registros vacíos de variables originales – Región Zona Sur

Nombre de Variable	Cantidad de registros sin datos	Porcentaje de registros sin datos respecto a total de registros en la región (%)
NOMBRE_PUENTE	0	0
CÓDIGO_PUENTE	0	0
CÓDIGO I3MOP	0	0
REGIÓN	0	0
PROVINCIA	0	0
COMUNA	0	0
TUICIÓN	0	0
CODIGO_RUTA	0	0
CLASE_RUTA	0	0
KM	0	0
Tramo	0	0
TMDA	1091	36
P-A	1091	36
P-C	1091	36
P-C2E	1091	36
P-SR	1091	36
P-R	1091	36
P-C+2E	1091	36
P-B	1091	36
Veloc_tramo	40	1
Tipología General	2	0
Tipología Estructural	17	1
Año de construcción	2465	82
Longitud total (m)	1	0
Ancho total (m)	2	0
Ancho de Calzada (m)	2093	70
Ancho de Pasillo (m)	2999	100
Número de tramos	72	2
Número de Cepas	411	14
Material cepas	171	6
Fundación cepas	939	31
Material estribos	79	3
Fundación estribos	2813	94

Material fundación estribos	1976	66
Número de Vigas	620	21
Vigas	611	20
Losa	279	9
Espesor de Losa (m)	793	26
Carpeta	61	2
Fecha última inspección principal	76	3
Calificación automática	76	3
Calificación final	76	3
Número total de deterioros	79	3
Índice Estructural	69	2
Índice Seguridad Vial	69	2
Índice Contención Sistema de Contención	69	2
Índice Contención Barandilla	69	2
Índice Durabilidad	69	2
Índice Daños Menores	69	2
Deterioro	90	3
Causa	205	7
Gravedad	90	3
Extensión	90	3
Medición	90	3
Unidad medición	92	3
Calificación automática de deterioro	152	5
Unidad	152	5
Código	728	24

Tabla A5.3 Categorías de deterioro

Deterioro	Categoría de deterioro
Depositación	Deterioro Estético
Grafiti	Deterioro Estético
Hongos	Deterioro Estético
Pudrición / ataque de insectos	Deterioro Estético
Restos metálicos de ejecución (clavos, barras)	Deterioro Estético
Vegetación	Deterioro Estético
Armadura vista	Deterioro Estructural Mayor
Cárcava	Deterioro Estructural Mayor
Corrosión	Deterioro Estructural Mayor
Desplazamiento	Deterioro Estructural Mayor
Desprendimientos de hormigón con armadura vista	Deterioro Estructural Mayor
Fisuras	Deterioro Estructural Mayor

Grietas (>5 mm)	Deterioro Estructural Mayor
Nido de piedras	Deterioro Estructural Mayor
Rotura	Deterioro Estructural Mayor
Socavación local	Deterioro Estructural Mayor
Deformación	Deterioro Estructural Menor
Desprendimientos de hormigón sin armadura vista	Deterioro Estructural Menor
Falta de alineación	Deterioro Estructural Menor
Fisuras en mapa o en retícula	Deterioro Estructural Menor
Pátina, mancha de óxido	Deterioro Estructural Menor
Perdida de finos	Deterioro Estructural Menor
Pérdida de pieza	Deterioro Estructural Menor
Pérdida de tornillos, roblones, anclajes	Deterioro Estructural Menor
Piezas sueltas	Deterioro Estructural Menor
Alteración superficial	Deterioro No estructural
Baches, Ahuellamiento	Deterioro No estructural
Circulación o presencia de agua	Deterioro No estructural
Despegue	Deterioro No estructural
Eflorescencias	Deterioro No estructural
Humedades, filtraciones	Deterioro No estructural
Hundimiento	Deterioro No estructural
Lajación	Deterioro No estructural
Obstrucción	Deterioro No estructural
Pérdida de tratamiento protector	Deterioro No estructural

ANEXO 6.1 Ranking de influencia de variables

Variable	RelieFF	XGBoost	Random Forest
Clase de ruta	0.2472	0.0036	0.0078
TMDA	0.0834	0.0044	0.0035
TMDA de vehículos livianos	0.0753	0.0000	0.0068
TMDA de vehículos pesados	0.0977	0.0521	0.0054
Proporción de vehículos livianos	0.0852	0.0139	0.0059
Proporción de vehículos pesados	0.0852	0.0000	0.0054
Tipología estructural	0.2202	0.0000	0.0034
Longitud total (m)	0.0442	0.0116	0.0104
Ancho total (m)	0.0926	0.0560	0.0091
Número de tramos	0.0605	0.0093	0.0047
Número de cepas	0.1650	0.0002	0.0021
Material de cepas	0.0811	0.0000	0.0131
Número de vigas	0.1131	0.0023	0.0048
Vigas	0.1457	0.0185	0.0227
Número total de deterioros	0.1376	0.2260	0.0246
Deterioros por longitud de puente	0.0902	0.0496	0.0122
Deterioros por superficie de tablero	0.0790	0.0157	0.0102
Causa	0.0191	0.0048	0.0062
Unidad	0.0076	0.0046	0.0042

Figura A6.1.1 Ranking de Variables Filtradas (Clasificación) - Región Zona Centro

Variable	Relieff	XGBoost	Random Forest
Clase de ruta	0.1879	0.0119	0.0013
TMDA	0.0911	0.0012	0.0046
TMDA de vehículos livianos	0.1019	0.0020	0.0037
TMDA de vehículos pesados	0.0704	0.0000	0.0035
Proporción de vehículos livianos	0.1281	0.0249	0.0044
Proporción de vehículos pesados	0.1281	0.0000	0.0046
Tipología general	0.0003	0.0098	-0.0001
Longitud total (m)	0.1024	0.0565	0.0049
Ancho total (m)	0.0576	0.0567	0.0027
Material de cepas	0.1431	0.0000	0.0039
Número de vigas	0.1285	0.0035	0.0022
Vigas	0.0750	0.0000	0.0028
Losa	0.0766	0.0001	0.0109
Espesor de losa (m)	0.1461	0.0265	0.0110
Carpeta	0.1764	0.0069	0.0112
Material	0.0585	0.0000	0.0011
Número total de deterioros	0.1774	0.0587	0.0146
Deterioros por longitud de puente	0.1612	0.0353	0.0084
Deterioros por superficie de tablero	0.1173	0.0252	0.0043
Unidad	-0.1633	0.0004	0.0019

Figura A6.1.2 Ranking de Variables Filtradas (Clasificación) - Región Zona Sur

Variable	XGBoost	Random Forest
TMDA	0.0978	0.1704
TMDA de vehículos pesados	0.0249	0.0718
Longitud total (m)	0.0746	0.3607
Ancho total (m)	0.5031	0.9816
Material de cepas	0.1454	0.0241
Vigas	0.0945	0.0478
Espesor de losa (m)	0.2303	0.0625
Número total de deterioros	0.5472	0.8518
Deterioros por longitud de puente	0.1140	0.2475

Figura A6.1.3 Ranking de Variables Filtradas (Regresión) - Región Zona Centro

Variable	XGBoost	Random Forest
TMDA	0.0904	0.1494
Proporción de vehículos livianos	0.1200	0.1925
Longitud total (m)	0.3368	0.1808
Ancho total (m)	0.0802	0.1516
Material de cepas	0.0820	0.0679
Número de vigas	0.1081	0.1585
Espesor de losa (m)	0.0733	0.0692
Carpeta	0.0882	0.1026
Número total de deterioros	0.4992	0.5830
Deterioros por longitud de puente	0.1446	0.3571
Deterioros por superficie de tablero	0.0945	0.2041

Figura A6.1.4 Ranking de Variables Filtradas (Regresión) - Región Zona Sur

**UNIVERSIDAD DE CONCEPCIÓN – FACULTAD DE INGENIERÍA
RESUMEN DE MEMORIA DE TÍTULO**

Departamento	: Departamento de Ingeniería Civil
Carrera	: Ingeniería Civil
Nombre del memorista	: Lorenzo Reyes Pedernera
Título de la memoria	: Identificación de variables influyentes en la calificación de condición de deterioro de puentes usando algoritmos de Machine Learning.
Fecha de la presentación oral	:
Profesor(es) Guía	: Tomás Benjamín Echaveguren Navarro
Profesor(es) Revisor(es)	: Juan Antonio Carrasco Montagna, Víctor Hugo Aguilar Vidal
Concepto	:
Calificación	:

Resumen

Se tiene como propósito caracterizar la influencia de variables explicativas en la calificación de condición de puentes en bases de datos en Chile. Se mide la influencia de variables explicativas utilizando herramientas de Machine Learning (ML). Se utilizó la técnica de Importancia de Atributos por Permutación (aplicada en Random Forest y Extreme Gradient Boosting) y la técnica ReliefF. Se destacan las variables más relevantes en la calificación de condición de un total de 58 variables presentes en bases de datos de dos regiones del centro y sur de Chile.

Las características de mayor influencia en la calificación de condición de puentes fueron el número de deterioros, el ancho total de puentes, la materialidad de la losa, la clase de ruta, la cantidad de deterioros por longitud de puente, y la materialidad de carpeta. De las variables influyentes solo la cantidad de deterioros y la causa de deterioros se encuentran incluidas en la calificación de condición actual. Este ranking de variables influyentes en la calificación de condición de puentes puede orientar la calibración de modelos de comportamiento de puentes.