Aplicación de la regresión logística para la predicción de roturas de tuberías en redes de abastecimiento de agua

Robles-Velasco A, Cortés P, Muñuzuri J, Barbadilla-Martín E

Recibido: 29 de Octubre de 2019
Aceptado: 15 de Noviembre de 2019

https://doi.org/10.37610/dyo.v0i70.570

Resumen

Las roturas de tuberías en redes de abastecimiento de agua provocan serios problemas para las compañías encargadas de su gestión. Con objeto de reducir el número de roturas inesperadas, se propone un método predictivo de clasificación de las tuberías que utiliza la regresión logística, junto con técnicas avanzadas de procesamiento de datos, como el equilibrado de clases o la validación cruzada.

La metodología se ha aplicado al caso real de la red de abastecimiento de Sevilla. Los resultados muestran que podría llegar a predecirse el 85.9% de las roturas de tuberías, siendo 76.6% la precisión total del modelo.

Palabras clave

Regresión logística; Red de abastecimiento de agua; Predicción de roturas de tuberías; Aprendizaje automático

1. Introducción

El deterioro de una red de abastecimiento de agua tiene como síntomas la aparición de roturas frecuentes (Pelletier, Mailhot, & Villeneuve, 2003). Son muchos los estudios que tratan de determinar las causas de las roturas de los distintos elementos que constituyen el sistema, así como sus modos de fallo. Actualmente, las empresas encargadas de la gestión de estas infraestructuras están enfrentando grandes retos en el mantenimiento y la sustitución de las mismas. Una rotura inesperada, además de causar daños materiales, supone la interrupción del suministro, lo que deriva en la disminución de la calidad del servicio. Por consiguiente, un sistema robusto de predicción de roturas provocaría una mejora del servicio y un ahorro de costes significativo.

En este estudio se utiliza un modelo de regresión logística como sistema predictivo de roturas de tuberías. Con ello, se pretende predecir el comportamiento futuro de las redes de abastecimiento de agua en base a su histórico de datos. En la sección 2 se incluye un estado del arte de las técnicas aplicadas hasta el momento para este mismo propósito. A continuación, en la sección 3 se explica la metodología, es decir, la regresión logística. Con objeto de validar dicha metodología, se hace uso de los datos de la red de abastecimiento de Sevilla. La sección 4 incluye la descripción de la red y el análisis de los resultados obtenidos. Por último, las conclusiones son expuestas en la sección 5.

✉ Robles-Velasco A * , **
arobles2@us.es

Cortés P *
Muñuzuri J *
Barbadilla-Martín E *
* Dpto. de Organización Industrial y Gestión de Empresas II. ETSI. Universidad de Sevilla. C/ Camino de los Descubrimientos S/N, 41092 Sevilla (Spain)
** Cátedra del Agua (EMASESA-Universidad de Sevilla)

2. Revisión bibliográfica

Son muchas las variables que pueden influir en la rotura de una tubería, desde propiedades físicas de la misma, como el diámetro o el material, hasta condiciones climatológicas o externas a la red. La actual tendencia a almacenar grandes cantidades de datos, ha hecho posible el estudio en mayor profundidad de estas variables, sus interacciones y su relación con la rotura. Sin embargo, no siempre se dispone de todas las variables que influyen en la rotura, por lo que los investigadores se enfrentan al desafío de conseguir predicciones precisas con información limitada.

Existen dos posibles enfoques del problema. El primero es el análisis descriptivo mediante el cual se estudian las causas de las roturas y las posibles situaciones que las potencian. El segundo es el análisis predictivo, objeto de estudio de este trabajo, cuyo objetivo final es la predicción de las roturas para así poder evitarlas, disminuyendo las consecuencias negativas que éstas implican. Según Amaitik y Buckingham (2018), existen tres tipos de análisis predictivos: estadísticos, probabilísticos e inteligencia artificial.

Entre los modelos estadísticos más utilizados en la predicción de roturas de tuberías, juegan un papel importante los modelos de supervivencia (Debón, Carrión, Cabrera, &

Solano, 2010; Kabir, Tesfamariam, & Sadiq, 2015; Yamijala, Guikema, & Brumbelow, 2009). Por lo general, estos modelos predicen el tiempo hasta la ocurrencia de cierto suceso de interés. Aunque los resultados son satisfactorios, estos modelos no permiten contemplar las tuberías que no sufren roturas, lo que disminuye su aplicabilidad a casos reales.

Entre los métodos probabilísticos más destacados están las redes bayesianas, que son grafos dónde los nodos representan las variables y los arcos las relaciones probabilísticas entre ellas. Esta probabilidad condicionada, en la mayoría de los casos, se obtiene de opiniones de expertos (Kabir, Tesfamariam, Francisque, & Sadiq, 2015), lo que puede originar errores. Las redes bayesianas son generalmente referidas como sistemas inteligentes por el modo en que la información se transmite por ellas (Royce, Seth, & Henneman, 2014).

Las redes neuronales (ANN) conforman uno de los mayores referentes de la inteligencia artificial. En (Jafar, Shahrour, & Juran, 2010), se compara la capacidad predictiva de varias redes con diferentes configuraciones. En este trabajo, se reconoce la influencia de las roturas previas de tuberías en la aparición de nuevas roturas. Otros estudios defienden que los algoritmos de máquinas de vector soporte (Support Vector Machine) son preferibles a las ANN para trabajar con sistemas de distribución de agua, ya que logran reproducir mejor los patrones de comportamiento físico (Shirzad, Tabesh, & Farmani, 2014).

Además de los métodos mencionados, también existen en la literatura muchos trabajos que aplican lógica difusa y técnicas multicriterio como el Analytic Hierarchy Process AHP (Al-Zahrani, Abo-Monasar, & Sadiq, 2016; Amaitik & Buckingham, 2018). La lógica difusa es el método más indicado cuando se trabaja con información imprecisa o ambigua. Otros autores usan fuzzy-clustering para dividir el total de tuberías en familias, demostrándose que después de esta división se consiguen mejores predicciones (Aydogdu & Firat, 2015).

La regresión logística, método escogido en este trabajo, se presenta como una alternativa a los modelos estadísticos nombrados anteriormente que permite contemplar tanto las tuberías que sufren alguna rotura, como las que no. Una de las ventajas de este método es que proporciona una salida que puede interpretarse como una probabilidad. Esto hace que su implantación en la industria sea mucho más factible.

Varios autores ya han aplicado la regresión logística para predecir roturas de tuberías en redes de abastecimiento de agua, obteniendo muy buenos resultados (Debón et al., 2010; Wang, Dong, Wang, Tang, & Yao, 2013; Yamijala et al., 2009). En nuestro estudio, se emplean variables diferentes para explicar la rotura de tuberías. Además, se utiliza la matriz de confusión como métrica de calidad para medir la precisión de los resultados. Esta matriz permite identificar el porcentaje de roturas correctamente predichas aportando claridad a los resultados y permitiendo demostrar la utilidad de la técnica al caso de estudio.

3. Metodología

Esta sección está dividida en dos subsecciones. En primer lugar, se desarrolla el modelo de regresión logística utilizado como sistema predictivo. A continuación, se explican las principales técnicas que deben emplearse para procesar los datos de una red de abastecimiento de agua. Por último, se describen las métricas de calidad utilizadas para analizar los resultados.

3.1. Regresión logística

La regresión logística (Cox & Snell, 1989) es un método estadístico que permite modelar fenómenos cuya variable respuesta es cualitativa, estableciendo la probabilidad de pertenecer a una clase como una función de distribución logística (ec.1). Ésta es una función sigmoidal que genera una relación no lineal entre las variables.

$$p_i = \frac{1}{1 + e^{-wx_i}} \tag{1}$$

Resolver el modelo implica estimar un vector de pesos, w, asociado a las variables explicativas, xi, que maximizan la función de verosimilitud (ec.2). Esto equivale a minimizar la desviación total, asignándoles una probabilidad alta (cercana a 1) a las tuberías que se rompen y una probabilidad baja (cercana a 0) a las que no. El subíndice i se refiere a cada una de las observaciones que forman la muestra N.

$$l = \prod_{i=1}^{N} p_i^{y_i} (1 - p_i)^{1-y_i} \tag{2}$$

La salida del modelo es una variable binaria, yi, que representa si el suceso de interés ocurre o no. En este estudio, el suceso de interés es la rotura de una tubería. La ecuación (3) presenta el cálculo de dicha salida. En general, el umbral de riesgo se establece en 0.5, no obstante, puede modificarse en función de las exigencias del problema.

$$y_i = \begin{cases} 0 \; si \; p_i < umbral \; de \; riesgo \\ 1 \; si \; p_i \geq umbral \; de \; riesgo \end{cases} \tag{3}$$

3.2. Procesamiento de los datos

El preprocesamiento de los datos tiene un papel fundamental en el rendimiento y la calidad de los sistemas predictivos. Siempre que se trabaja con grandes cantidades de datos es necesario realizar una visualización inicial de éstos para analizar sus principales características y detectar posibles anomalías.

Los datos procedentes de redes de abastecimiento de agua suelen presentar variables categóricas. Estas variables deben ser transformadas en variables numéricas. En este estudio, se propone asignar un número a cada categoría en función de su tasa de rotura por unidad de longitud. Respecto a los huecos, del inglés missing values, se implementa un sistema de relleno de los mismos con la mediana de la variable. Esto evita eliminar observaciones por la falta de alguna de sus variables.

En sistemas complejos es común encontrar variables que se mueven en diferentes escalas. La transformación logarítmica es una opción eficaz para reducir la escala de aquellas variables que se mueven en órdenes de magnitud superiores al resto, unificando así la escala de las variables.

Otra característica, presente en el histórico de datos de redes de abastecimiento de agua, es que la rotura de tuberías es una variable totalmente desequilibrada. Esto ocurre porque el ratio de tuberías que rompen en un determinado periodo de tiempo es del orden de 1:100 con respecto al número de tuberías que no sufren ninguna rotura (Wang et al., 2013). Para evitar que esta característica de los datos deteriore la precisión de las predicciones, se propone utilizar técnicas de remuestreo en aquellos datos que se utilizan para estimar el modelo predictivo.

3.3. Métricas de calidad

La matriz de confusión (figura 1) se utiliza para medir la calidad de los resultados. Esta matriz enfrenta las predicciones con los datos reales, permitiendo analizar de forma sencilla los resultados de un sistema predictivo de clasificación. Por un lado, los verdaderos positivos o TP, del inglés True-Positive, son el número de predicciones correctas de clase 1. Por otro lado, los falsos negativos o FN, del inglés False-Negative, representan el número de predicciones erróneas de clase 1, lo que se traduciría como el número de roturas reales no predichas.

Figura 1 Matriz de confusión

Predicción \ Realidad	1	0
1	TP	FP
0	FN	TN

En concreto, se utilizan dos indicadores derivados de esta matriz: accuracy y recall. El primero recoge el porcentaje total de predicciones correctas y el segundo el porcentaje de roturas predichas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

La curva ROC, del inglés Receiver Operating Curve, es otra métrica que permite analizar la capacidad de un clasificador para evitar clasificaciones falsas. El área bajo la curva o AUC, del inglés Area Under the Curve, puede tomar valores entre 0 y 1. Un AUC igual a 0.5 correspondería a un clasificador aleatorio y, cuanto más cercano a 1, mejor será el clasificador.

4. Caso de estudio

Con objeto de validar la aplicabilidad de esta metodología al problema de estudio, se hace uso de los datos correspondientes a la red de abastecimiento de Sevilla. Esta red está formada por más de 3,800 kilómetros de tuberías y abastece a una población de más de 1 millón de personas.

4.1. Descripción y procesamiento de los datos

Las variables de entrada al modelo y sus respectivas unidades de medida se recogen en la tabla 1. La salida del modelo es una variable binaria $y_i\{0,1\}$, donde 1 representa que la tubería se rompe y 0 lo contrario.

Tabla 1 Descripción de las variables explicativas del modelo

Variable	Definición	Unidad	Media	Min	Max
Mat	Material de la tubería	-	-	-	-
Dia	Diámetro de la tubería	mm	152.18	20	1700
Edad	Edad de la tubería	Años	25.72	0	118
L	Longitud de la tubería	m	57.12	0.50	2522
Ac	Número de acometidas	-	2.08	0	71
T_red	Tipo de red (transporte o secundaria)	-	-	-	-
NOPF[1]	Número de roturas previas	-	0.04	0	10
F_pre	Variación de la presión	m	2.87	0	60.19

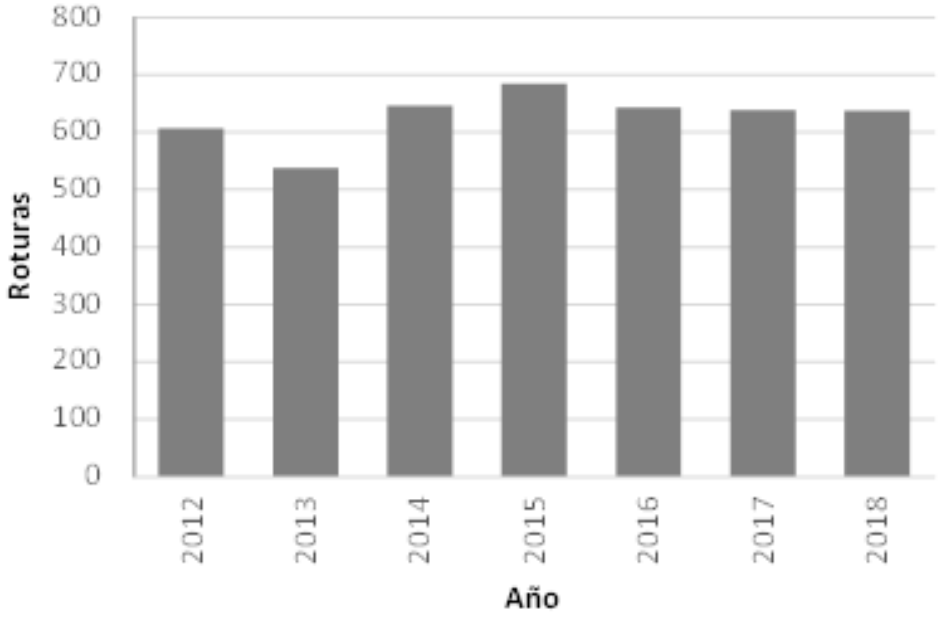
[1] Del inglés Number Of Previous Failures.

Para conseguir un funcionamiento eficiente del algoritmo, es primordial un buen preprocesamiento de los datos. En primer lugar, se ha procedido a limpiar los datos, eliminando ciertas variables por falta de fiabilidad. A continuación, a cada categoría de las dos variables categóricas, material y tipo de red, se le asigna un número en función de su tasa de rotura por unidad de longitud. La red de abastecimiento analizada se compone de tuberías de catorce materiales distintos. El material con mayor tasa de rotura es el hierro fundido (HF), mostrando una media de 4.3 roturas por kilómetro de tubería en los siete años de estudio. Mientras que la tasa de rotura de las tuberías de fundición dúctil (FD) es de 0.1, siendo uno de los materiales con menor número de roturas por kilómetro de tubería. En cuanto al tipo de red, la red secundaria presenta una tasa de rotura mayor a la red de transporte, 1.1 frente a 0.4.

Como se ha mencionado anteriormente, los huecos son rellenados con la mediana de cada variable. Además, se decide utilizar el logaritmo de las variables Dia y L en lugar de sus valores originales, para evitar que las variables se extiendan sobre varios órdenes de magnitud. Por último, todas las variables fueron estandarizadas.

El histórico de datos disponible consta de siete años y un total de 4,393 roturas. El total de tramos que componen la red varía entre 83,000 y 88,000 en los diferentes años de estudio. El número de roturas registradas en cada año se muestra en la figura 2, siendo este número siempre menor a 700 roturas. Esto supone que la muestra está totalmente desequilibrada, siendo el número de tramos de la clase mayoritaria, tuberías que no rompen, más del 99% del total de la muestra en cada año. Entrenar un modelo de regresión logística con una muestra desequilibrada resulta en priorizar la clasificación correcta de la clase mayoritaria. Como el objetivo principal de este estudio es predecir las roturas, clase minoritaria, se hace uso de una técnica de remuestreo (under-sampling) equilibrando las clases en el conjunto de entrenamiento. El conjunto de evaluación no se altera, representando así fielmente la realidad.

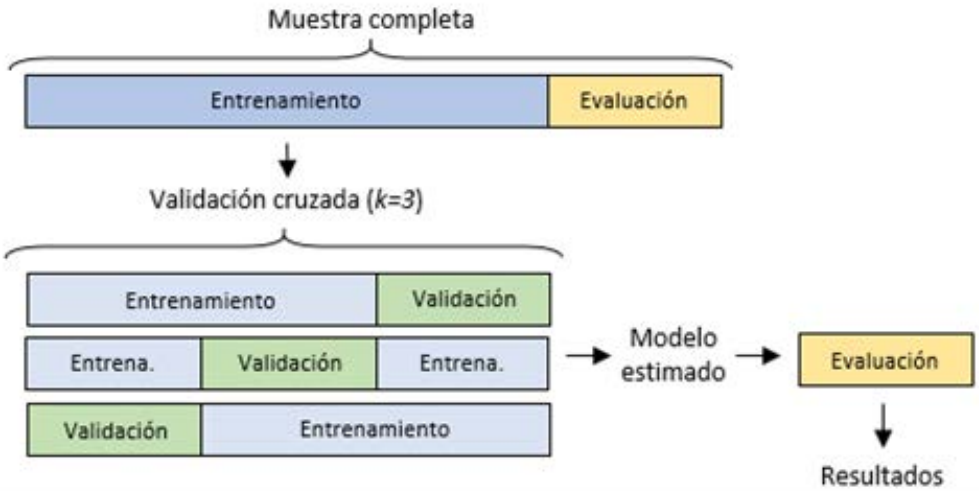
Figura 2 Roturas registradas en los años de estudio



4.2. Resultados

El lenguaje de programación utilizado para resolver el modelo es Python y, en concreto, la librería Scikit-learn (Pedregosa et al., 2011). La muestra, compuesta por un histórico de siete años, se divide aleatoriamente en dos partes, cinco años para el entrenamiento del modelo, y dos para su evaluación. Con el objetivo de obtener un modelo más generalizado, se aplica la validación cruzada, del inglés cross-validation, dividiendo el conjunto de entrenamiento a su vez en cinco partes (k=5). Este procedimiento permite aprovechar mejor los datos y obtener un modelo final más generalizado. En la figura 3 puede observarse una representación esque-mática de este mecanismo para k=3.

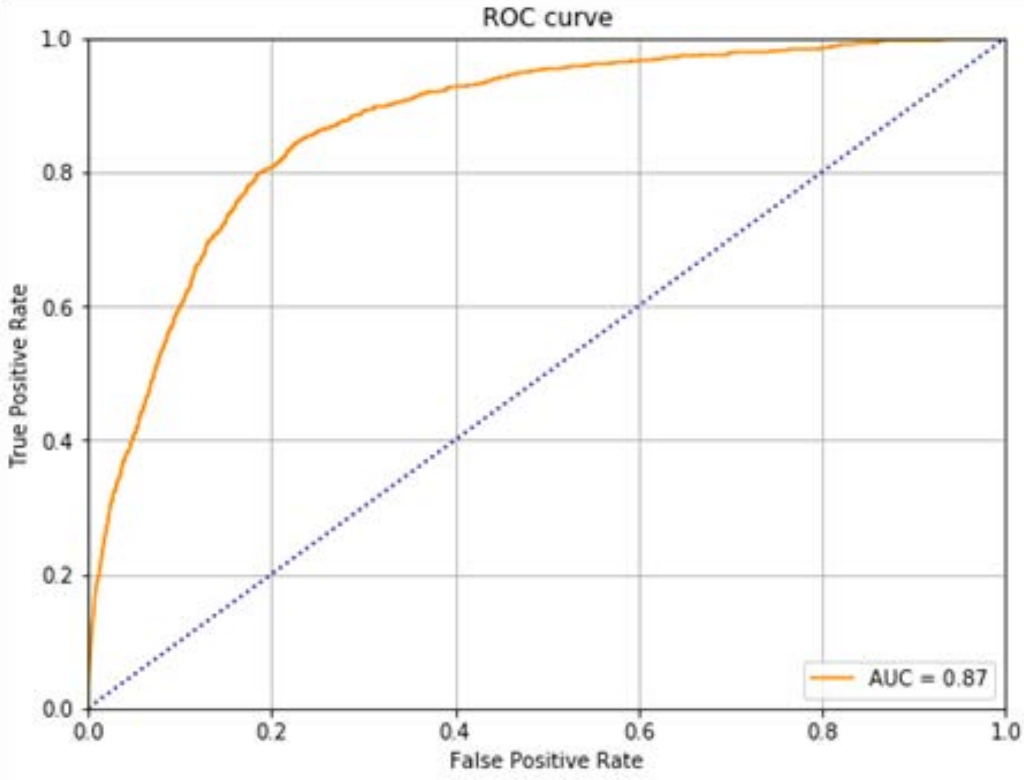
Figura 3 Ejemplo de validación cruzada para k=3



Tras estimar el modelo con los datos de entrenamiento, se consigue una precisión (accuracy) del 76.6% en los datos de evaluación. Esto significa que la condición del 76.6% de las tuberías es correctamente predicha. Se ha dado prioridad a predecir las roturas equilibrando las clases en la fase de entrenamiento, mientras los datos de evaluación mantienen su estructura y forma original, representando fielmente la realidad. Con ello, de entre todas las tuberías que se rompen en el conjunto de evaluación (1,054), un 85.9% (recall) son bien predichas, es decir, podrían haberse evitado 905 roturas.

La figura 4 muestra la curva ROC de los datos de evaluación. Se observa que el AUC obtenido es muy alto, 0.87. Cuanto más cercano a 1, mejor es la capacidad discriminante del clasificador. Por ello, se demuestra que utilizando el modelo de regresión logística junto con el procesamiento previo de los datos se consiguen predicciones de gran calidad.

Figura 4 Curva ROC de los datos de evaluación



Los pesos estimados del modelo (tabla 2) permiten extraer información interesante acerca de las variables y su relación con la rotura.

Figura 4 Curva ROC de los datos de evaluación

w_0	w_1 Mat	w_2 Log(Dia)	w_3 Edad	w_4 Log(L)	w_5 Ac	w_6 T_red	w_7 NOPF	w_8 F_Pre
-1.183	0.978	-0.152	0.239	0.917	0.000	-0.063	0.270	-0.103

Mat: Material; Log(Dia): Logaritmo del diámetro; Ac: Número de acometidas; Log(L): Logaritmo de la longitud; T_red: Tipo de red; NOPF: Número de roturas previas; F_Pre: Variación de la presión

La probabilidad de rotura de una tubería cuando todas las variables son iguales a sus medias es 0.23, lo que se extrae de sustituir w0 en la expresión (4).

$$p = \frac{1}{1 + e^{-w_0}} \quad (4)$$

Los demás pesos representan en qué grado de magnitud varían las probabilidades de rotura por unidad de cambio de las variables. Se puede deducir, por tanto, que el material y el logaritmo de la longitud son las variables más influyentes, mientras que las acometidas y el tipo de red apenas participan en la discriminación por clases. El signo de estos pesos también indica si la relación entre las variables y la rotura es directa o inversa. Por ejemplo, se observa como las tuberías con diámetros más pequeños son más propensas a romper.

5. Conclusiones

En este estudio, se aborda la problemática referente a la aparición de roturas inesperadas en las tuberías de las redes de abastecimiento de agua. Estas roturas causan daños materiales y, en muchas ocasiones, el corte del suministro. La regresión logística es un método estadístico que permite predecir las roturas mediante la clasificación de las tuberías en dos clases: (i) tuberías más propensas a la rotura; y (ii) tuberías con menos propensión a la rotura. Para estimar el modelo, calculando los pesos asociados a las variables explicativas, se necesitan datos de entrenamiento. Posteriormente, dichos pesos estimados se utilizan para predecir la clase de nuevas muestras.

Esta técnica se ha aplicado a la red de abastecimiento de Sevilla, utilizando un histórico de roturas de siete años. Las características de dicha red, como su edad y el material de sus tuberías, son comunes a las del resto de redes de abastecimiento de España y parte de Europa. Por otro lado, su tamaño, 3,800 kilómetros de tuberías, representa el de una ciudad de tamaño medio o grande, de aproximadamente 1 millón de habitantes. Los datos utilizados han sido obtenidos de un sistema de información geográfica (GIS), que actualmente es utilizado por la mayoría de las compañías que gestionan estas infraestructuras. Por ello, los resultados alcanzados en este estudio pueden extrapolarse a otras redes de abastecimiento que compartan estas características.

Previo a la estimación del modelo, los datos disponibles han sido procesados aplicando ciertas transformaciones de los mismos como la transformación logarítmica o la estandarización. Una de las características principales de los datos de redes de abastecimiento de agua es la existencia de muchas más tuberías que no se rompen frente a las tuberías que sí lo hacen, siendo éste un problema de clases desequilibradas. Entrenar un modelo con una muestra desequilibrada supone obtener predicciones mucho más precisas de la clase mayoritaria. Por lo tanto, el entrenamiento del modelo se debe realizar con muestras equilibradas mediante una técnica de muestro. Además, se han incorporado mecanismos de machine learning, como la validación cruzada o el relleno de huecos, consiguiendo que el rendimiento del modelo mejorara sustancialmente.

El porcentaje de roturas predichas en el conjunto de evaluación, histórico de dos años, es del 85.9% (accuracy), demostrándose así la eficacia de esta técnica. Las variables que se han identificado como las más influyentes en la rotura son el material y la longitud de la tubería, seguido por el número de roturas previas y la edad. Por otro lado, se observa que las tuberías de menor diámetro son más propensas a la rotura.

La regresión logística se presenta como una opción eficaz y sencilla de aplicar. En este estudio se demuestra su aplicabilidad a redes de abastecimiento de gran tamaño. La interpretación de los resultados es directa, ya que se obtiene la probabilidad de rotura asociada a cada tubería. Esto hace que su integración en la industria sea mucho más factible. Una de las desventajas de este método es que no evalúa la evolución de las variables en el tiempo (Wilson, Filion, & Moore, 2017). Futuras líneas de investigación deberían explorar esta carencia. Además, se propone la aplicación de otras técnicas probabilísticas y de inteligencia artificial para predecir roturas de tuberías, con objeto de identificar cual es la que mejor se ajusta al problema en cuestión.

Agradecimientos

Los autores desean agradecer el apoyo y la financiación de la Catedra del Agua, proyecto conjunto de EMASESA (Empresa de Abastecimiento y Saneamiento de Aguas de Sevilla y su área metropolitana) y la Universidad de Sevilla (VI PPIT-US), a través de un programa de doctorado industrial.

Este trabajo ha sido presentado en el 13th International Conference on Industrial Engineering and Industrial Management, XXIII Congreso de Ingeniería de Organización celebrado en Gijón, España, los días 11 y 12 de julio de 2019.

6. Referencias

Al-Zahrani, M., Abo-Monasar, A., & Sadiq, R. (2016). Risk-based prioritization of water main failure using fuzzy synthetic evaluation technique. Journal of Water Supply: Research and Technology - AQUA, 65(2), 145–161. https://doi.org/10.2166/aqua.2015.051

Amaitik, N. M., & Buckingham, C. D. (2018). Developing a hierarchical fuzzy rule-based model with weighted linguistic rules: A case study of water pipes condition prediction. Proceedings of Computing Conference 2017, 2018-Janua(July), 30–40. https://doi.org/10.1109/SAI.2017.8252078

Aydogdu, M., & Firat, M. (2015). Estimation of Failure Rate in Water Distribution Network Using Fuzzy Clustering and LS-SVM Methods. Water Resources Management, 29(5), 1575–1590. https://doi.org/10.1007/s11269-014-0895-5

Cox, D. R., & Snell, E. J. (1989). Analysis of Binary Data (2nd ed.). London: Chapman and Hall Ltd.

Debón, A., Carrión, A., Cabrera, E., & Solano, H. (2010). Comparing risk of failure models in water supply networks using ROC curves. Reliability Engineering and System Safety, 95(1), 43–48. https://doi.org/10.1016/j.ress.2009.07.004

Jafar, R., Shahrour, I., & Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. Mathematical and Computer Modelling, 51(9–10), 1170–1180. https://doi.org/10.1016/j.mcm.2009.12.033

Kabir, G., Tesfamariam, S., Francisque, A., & Sadiq, R. (2015). Evaluating risk of water mains failure using a Bayesian belief network model. European Journal of Operational Research, 240(1), 220–234. https://doi.org/10.1016/j.ejor.2014.06.033

Kabir, G., Tesfamariam, S., & Sadiq, R. (2015). Predicting water main failures using Bayesian model averaging and survival modelling approach. Reliability Engineering and System Safety, 142, 498–514. https://doi.org/10.1016/j.ress.2015.06.011

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Olivier, G., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2

Pelletier, G. V., Mailhot, A., & Villeneuve, J.-P. (2003). Modeling Water Pipe Breaks-Three Case Studies. Journal of Water Resources Planning and Management, 129, 115–123.

Royce, A. F., Seth, D. G., & Henneman, L. (2014). Bayesian Belief Networks for predicting drinking water distribution system pipe breaks. Reliability Engineering and System Safety, 130, 1–11. https://doi.org/10.1016/j.ress.2014.04.024

Shirzad, A., Tabesh, M., & Farmani, R. (2014). A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. KSCE Journal of Civil Engineering, 18(4), 941–948. https://doi.org/10.1007/s12205-014-0537-8

Wang, R., Dong, W., Wang, Y., Tang, K., & Yao, X. (2013). Pipe failure prediction: A data mining method. Proceedings - International Conference on Data Engineering, 1208–1218. https://doi.org/10.1109/ICDE.2013.6544910

Wilson, D., Filion, Y., & Moore, I. (2017). State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. Urban Water Journal, 14(2), 173–184. https://doi.org/10.1080/1573062X.2015.1080848

Yamijala, S., Guikema, S. D., & Brumbelow, K. (2009). Statistical models for the analysis of water distribution system pipe break data. Reliability Engineering and System Safety, 94(2), 282–293. https://doi.org/10.1016/j.ress.2008.03.011