

Inteligencia Artificial Aplicada al Método Backward Seismic Analysis

Artificial Intelligence Applied to the Backward Seismic Analysis Method

Möller-Acuña P ¹
Patricio Pineda Nalli ²

Resumen

Este trabajo presenta aplicaciones del método de Backward Seismic Analysis (BSA) para estanques de acero utilizando información de 382 estanques en operación durante grandes terremotos subductivos: Valdivia 1960, Chile Central 1985, Tocopilla 2007, El Maule 2010, además de Alaska (1964) y otros ocurridos en Estados Unidos entre 1933 y 1995 (subductivos y corticales). Se ha observado que la mayoría de los estanques sin anclajes han fallado durante grandes terremotos, siendo diseñados principalmente con los códigos API 650-E, AWWA-D100 y NZSEE, los cuales contienen procedimientos similares para estimar las solicitaciones sísmicas, pero con diferentes métodos de diseño. Para las conferencias de STESSA 2012, STESSA 2015, 16WCEE 2017, ACHISINA 2019, 17WCEE 2020, STESSA 2021 y el trabajo de tesis de Maestría en Ingeniería Sísmica del autor, se investigaron las causas de las fallas concluyendo que los estanques fueron diseñados principalmente con API 650-E no estaban anclados, además los códigos de diseño no consideran aspectos relevantes que condicionan la respuesta sísmica. En este trabajo desarrollamos un modelo de predicción basado en la información histórica ya descrita, capaz de predecir de manera eficiente si un estanque presentara fallas durante algún terremoto. Se evaluaron varios algoritmos y métodos de aprendizaje automático, encontramos que el método "Random Forest" exhibe los mejores resultados en general. Los resultados obtenidos en la predicción de fallas de estanques alcanzan más (90%) de eficiencia en la mayoría de los escenarios evaluados.

Palabras claves: Estanque, Anclado, Backward, Subducción, Aprendizaje Autonomico, Random Forest.

1 Artificial Intelligence Applied to the Backward Seismic Analysis Method

2 Dra. Patricia Möller-Acuña, Universidad Autónoma de Chile, <https://orcid.org/0000-0003-0980-7735>, patricia.moller@uautonoma.cl

3 Consultor en Ingeniería Sísmica_PPN Ingeniería, M.Sc. Universidad de Chile, patricio.pineda@ppning.com

Abstract

This work presents applications of the Backward Seismic Analysis (BSA) method for steel tanks using information from 382 tanks in operation during large subductive earthquakes: Valdivia 1960, Central Chile 1985, Tocopilla 2007, El Maule 2010, in addition to Alaska (1964) and others that occurred in the United States between 1933 and 1995 (subductive and cortical), it has been observed that most of the steel tanks unanchored have failed and were designed mainly with the codes API 650-E, AWWA-D100, and NZSEE, which contain similar procedures for estimating seismic forces, but with different design methods. For the conferences of STESSA 2012, STESSA 2015, 16WCEE 2017, ACHISINA 2019, 17WCEE 2020, STESSA 2021 and the work thesis of Master in Seismic Engineering of the author (PPN), the causes of the failures were investigated concluding that the tanks were designed mainly with API 650-E code and were unanchored, in addition the design codes do not consider relevant aspects that condition the seismic response. In this work we develop a prediction model based on the historical information already described, capable of efficiently predicting if a steel tanks will present failures during an earthquake. Various algorithms and machine learning methods were evaluated, we found that the "Random Forest" method exhibits the best results overall. The results obtained in the prediction of steel tanks failures reach more (90%) efficiency in most of the evaluated scenarios

Keywords: Tanks, Anchored, Backward, Subduction, Learning, Autonomic, Random, Forest

Introducción

Dentro de las ramas de la inteligencia artificial se encuentra el aprendizaje automático, que es un proceso a través del cual un computador es capaz de aprender de una experiencia para aplicarlo sobre una tarea de aprendizaje la cual es finalmente medida por la performance, según Mitchell [1]. Es así como de forma más específica se generan algoritmos que le permiten al computador generalizar comportamientos a partir de la experiencia proporcionada con el objetivo de repetir esta generalización en nuevos ejemplos.

Actualmente el diseño sísmico de estanques de acero se basa en las disposiciones de los códigos de diseño más utilizados en el mundo, tales como: API 650 en su apéndice E [2], AWWA-D100 [3] y NZSEE [4], a pesar de esto, se ha observado que la mayoría de los estanques sin anclajes han fallado durante grandes terremotos y fueron diseñados principalmente con el código API 650-E. Según lo anterior, es necesario revisar y modificar los criterios de diseño para el cálculo de los esfuerzos admisibles del manto, dado que contienen procedimientos similares para estimar las solicitaciones sísmicas, pero con diferentes métodos de diseño. En Chile se dispone de la norma chilena

NCh2369 y el anteproyecto de norma chilena en preparación por el ICHA (Instituto Chileno del Acero), este último se basa en el método Backward Seismic Analysis (BSA) [5, 6, 7, 8, 9, 10, 11, 12]. Con el método BSA es posible determinar los coeficientes sísmicos para el diseño de los estanques, considerando la respuesta sísmica real de una serie de estanques (382) en operación durante los terremotos de Valdivia en 1960, Chile Central en 1985, Tocopilla 2007, El Maule 2010, además de Alaska en 1964 y de los Estados Unidos entre los años 1933 y 1995.

En este artículo se presenta un sistema con un modelo de predicción de daño de estanque que utiliza información histórica de carácter internacional. Este modelo es capaz de predecir de manera eficiente y con hasta 93% de eficiencia esta predicción, además se entrega un análisis de importancia de cada uno de los atributos considerados para realizar los análisis de predicción.

Metodología

Por medio de Inteligencia Artificial se validará el método BSA consistente en evaluar el comportamiento sísmico de estanques de acero, considerando las características del estanque en operación durante el sismo: geometría general y espesores de planchas, altura de llenado, tipos de suelos de fundación, códigos de diseño utilizados, registros sísmicos, directividad sísmica, sin daños observados, con pandeo del manto y colapsos. El método BSA se basa en información real del comportamiento sísmico de estanques de acero, permitiendo clasificarlos según sus dimensiones y esbelteces, dentro de rangos seguros con daños menores, daños considerables y reparables hasta valores con alto riesgo de colapso. Para aplicar este método es necesaria la siguiente información de los estanques: dimensiones generales y espesores del manto, tipos de fundaciones y sistemas de anclajes, propiedades del suelo de fundación, tipo de líquido almacenado, niveles de llenado al momento del sismo, sismicidad y aceleraciones máximas del suelo de las zonas de emplazamiento del estanque, daños observados durante terremotos, características del techo: tipo (cónico, flotante, domo), estructuración, peso, criterios de diseño y códigos utilizados, planos de fabricación y As Built.

La clasificación supervisada es una de las tareas más frecuentes llevadas a cabo por IA. Puede ser a través de modelos desarrollados en función de la estadística (Regresión Logística, Análisis Discriminante) o bien por IA (Redes Neuronales, Inducción de Reglas, árboles de Decisión, Redes Bayesianas) los cuales son capaces de realizar las tareas propias de la clasificación [13,14]. Obteniendo de esta forma un modelo de clasificación el cual indicara si los estanques presentaran o no daño, con esta información se puede construir la matriz de confusión la cual permite visualizar el desempeño del modelo.

Caracterización de los datos

Se cuenta con 382 casos de estanques de los cuales 72 presentaron daños y 308 tuvieron buen comportamiento sísmico durante los terremotos considerados para los estudios.

1.- Pre-procesamiento de datos (Data preprocessing)

La limpieza de datos es la tarea fundamental de la Inteligencia Artificial [15].

Los datos de los estanques se analizaron para eliminar, corregir y estandarizar aquellos que presentaban anomalías en función del resto mediante algoritmos en lenguaje Python [16]

2.- Diseño de atributos (Feature Desing)

Se entiende por atributo al conjunto de mediciones de los estanques durante un proceso sísmico. Tales como:

Mag = Magnitud del sismo

G = Peso específico del líquido en el tanque

Hm(m) = Altura del manto tanque

HLL(m) = Altura del líquido en el tanque durante el sismo

D/H= relación de esbeltez del tanque

Cps = coeficiente sísmico obtenido del método BSA

ts = espesor de las planchas del manto del primer anillo inferior

De =diámetro del estanque

Fc = tensión admisible de compresión del manto del estanque

Vs = esfuerzo de corte sísmico

Mt = masa total del estanque

Ms = momento volcante sísmico

Fy = tensión de fluencia del acero

E = módulo de elasticidad de acero

Anclaje = El tanque se encuentra o no anclado

Daño = El tanque presento daño o no durante el sismo

3.- Selección de atributos (Feature Selection)

La selección de atributos permite reconocer las características principales de un conjunto de datos que están relacionadas con la variable de respuesta [17]. (¿Cuáles son las características que se relacionan con la presencia o no de daño?)

Chi-square test

Una prueba de chi-square se usa en estadísticas para probar la independencia de dos eventos. Con las variables se puede obtener el recuento observado O y el recuento esperado E [18]. chi-square mide como el recuento E y el recuento O se desvían entre sí (Ver formula 1)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Formula 1: Chi-square

Esta prueba se realizó a través de un algoritmo implementado en Python utilizando la librería Sklearn [19].

F-test

F-test es una prueba estadística que se utiliza para comparar entre modelos y verificar si la diferencia es significativa entre ellos. F-test hace un modelo de prueba de hipótesis X e Y.

Para evaluar se comparan errores de mínimos cuadrados en ambos modelos y se verifica si la diferencia en los errores entre el modelo X e Y es significativa o se introdujo de forma azarosa [17]. Para realizar este test se utilizó ANOVA a través de un algoritmo programado en Python utilizando la librería Sklearn [19].

4.- Balance de clases

Para realizar un balance de clases se utilizó la librería Sklearn de Python [19] específicamente la función resample (en una proporción de 70 – 30) el cual integra muestras sintéticas de la clase minoritaria utilizando diversos algoritmos que siguen la tendencia del grupo minoritario.

5.- Machine Learning

Dado el set de referencia anteriormente descrito, se evaluaron diversos métodos de clasificación como Naive Bayes [20], Random-Forest [21] y [22], Árboles de Decisión [13], SVM [23], Regresión logística [24], Gradient Boosting [25]. Todo los cuales fueron

implementados en el lenguaje de programación de Python específicamente utilizando la librería Sklearn[19].

6.- Entrenamiento y Testeo (Training and Testing)

Se realizaron 5 segmentaciones del set de datos y se entrenaron, parametrizaron y testearon los modelos de forma simultánea con el fin de encontrar el modelo óptimo para realizar la tarea designada a través de un algoritmo implementado en Python con el uso de la librería Sklearn

A partir de los resultados obtenidos, se obtuvieron los siguientes valores:

- VP son los *verdaderos positivos*: número de estanques que no fallo y el modelo indico que no fallo.
- FP son los *falsos positivos*: número de estanques que fallo y el modelo indico que no fallo.
- FN son los *falsos negativos*: número de estanques que no fallo y el modelo indico que fallo.
- VN son los *verdaderos negativos*: número de estanques que fallo y el modelo indico que fallo.
- EG *Eficiencia global*: corresponde a tasa de éxito general de la predicción.

De los cuales se derivan las siguientes medidas de performance definidas para este problema:

Matrices de confusión: Permite visualizar el desempeño de un algoritmo.

Curva Roc:

Realidad	Predicción	
	TN	FP
	FN	TP

Muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Esta curva representa dos parámetros: Tasa de verdaderos positivos (TP) y Tasa de falsos positivos (FP)

Precisión: Mide la calidad del modelo.

$$\text{precisión} = \text{TP} / (\text{TP} + \text{FP})$$

Recall: La métrica de exhaustividad informa sobre la cantidad de ejemplos que el modelo es capaz de identificar.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Resultados

Chi-square test

Se utilizó esta prueba para evaluar la dependencia del daño de los estanques con los atributos mencionados anteriormente. Se considera que existe una dependencia cuando el valor de la prueba es cercano a 0, por lo tanto, todos los atributos que muestran un valor superior son variables independientes que no presentan relación con el daño del tanque. (Imagen Resultados chi-square test)

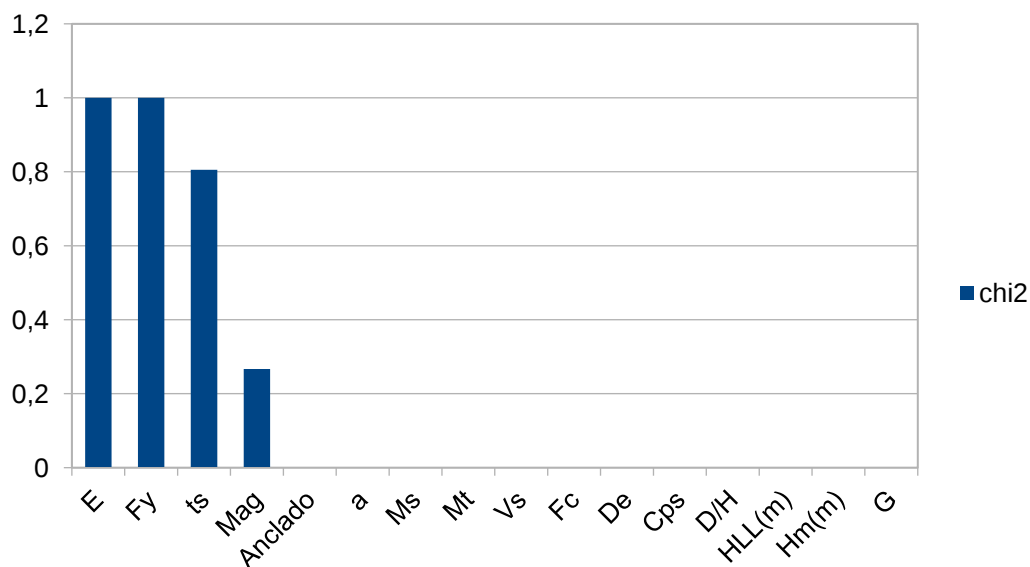


Figura1: Test Chi-square de los diferentes atributos analizados para predecir daño de estanques.

F-test (Anova)

Se utilizó esta prueba para medir la importancia de los atributos dependientes y estimar cuales son los más relevantes para estimar el posible daño en un tanque. Todos los atributos presentan un grado de relación con el resultado analizado (con o sin daño). Para realizar el modelo de clasificación se seleccionaron los 6 atributos que muestran los mayores puntajes en F-test, los cuales se observan la figura 2.

- * Anclado = 25,6
- * G = 14,08
- * Fc = 12,37
- * a = 10,23
- * Mt = 7,81
- * Vs = 7,13

Además, se graficó la precisión v/s recall para evaluar la efectividad del modelo mientras antes el recall llegue a 1 en función de la precisión mayor efectividad de clasificación presenta el modelo (Imagen Precision v/s Recall todos los atributos). El modelo de GBM (línea morada) presenta un mejor recall que el modelo de RF.

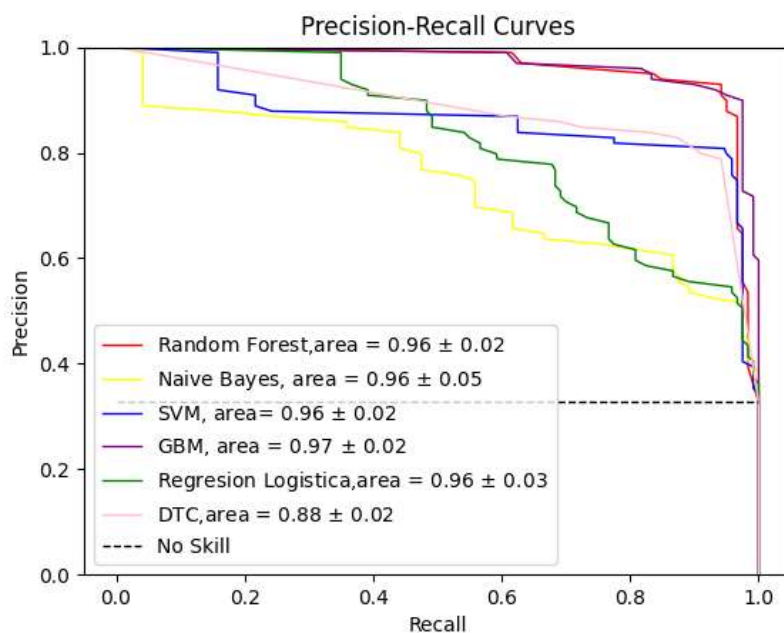


Figura 4: Curvas de Recall v/s Precision de los modelos de clasificación utilizando todos los atributos dependientes.

Atributos seleccionados por F-Test (ANOVA)

Al igual que el modelo de clasificación anterior los mejores modelos son RF y GBM con un 97% de precisión (Curva Roc atributos Anova). En comparación con el caso anterior al utilizar todos los atributos hay modelos como NB que disminuyen su precisión al reducir la cantidad de atributos.

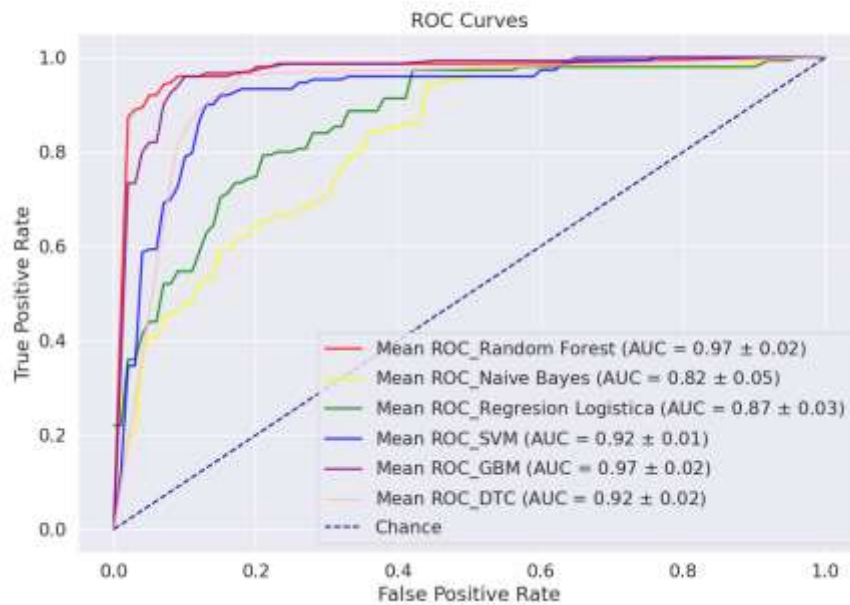


Figura 5: Curvas ROC de los modelos de clasificación utilizando los atributos seleccionados por F-Test.

El grafico de relación de precision v/s recall se puede observar que el modelo GBM presenta un mejor recall, pero RF mejora su recall en comparación con el uso de todos los atributos.

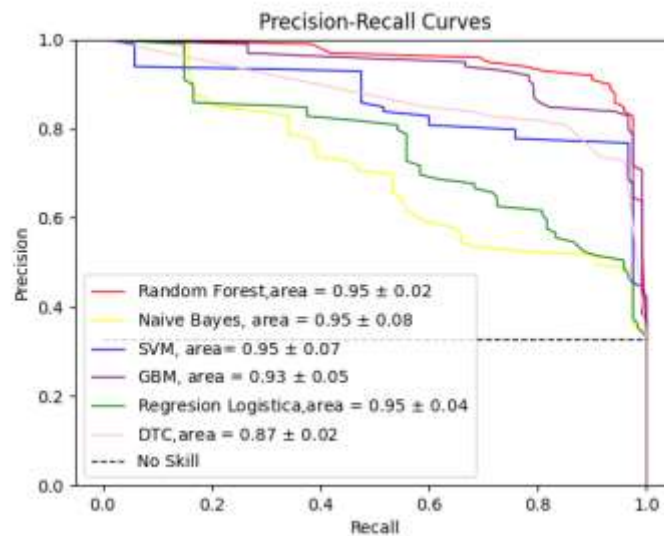


Figura 5: Curvas de Recall v/s Precision de los modelos de clasificación utilizando los atributos seleccionados por F-Test.

Conclusiones

El daño de los estanques registrados en datos históricos muestra efectos devastadores para el país donde esto ocurre. En este trabajo se presenta el desarrollo de un modelo de predicción de posible daño en los estanques basado en información de registro histórica y técnicas de aprendizaje automático. Pudiendo concluir que el análisis de datos y uso de herramientas de IA permite analizar cuáles son las características y/o factores más importantes para que un estanque presente o no daño, donde juega un rol clave el anclaje del estanque, ya que esta se presenta como la característica más relevante para que un estanque no presente daño, sin embargo, por sí solo no puede evitar que ocurra este fenómeno y es importante mencionar que la magnitud del sismo es una variable independiente del daño del tanque.

Diversos métodos de aprendizaje automático fueron evaluados para esta tarea encontrando que RF y GBM son los mejores modelos para realizar el proceso de clasificación de este problema (presenta o no daño) en los distintos escenarios y condiciones evaluadas para registros. Esto nos indica que estos algoritmos se adaptan mejor a las condiciones planteadas en la problemática propuesta en este trabajo asociada a “predecir de forma temprana la ocurrencia de daño de un estanque utilizando información histórica”. El modelo desarrollado alcanza una eficiencia del 93% de eficiencia. Utilizando el modelo de clasificación GBM se podría estimar si un tanque va o no a presentar daño durante un sismo, en función de su anclaje, G , a , F_s , M_t , V_s . Este trabajo por tanto entrega resultados concluyentes en cuanto a que es si posible determinar los principales atributos que juegan un rol clave en el modelo recién mencionado.

Finalmente se determina los resultados con el bagaje teórico que antecede y sustenta al trabajo desarrollado. Supone una muestra de las reflexiones del autor con relación al nuevo conocimiento generado en dialogo con el conocimiento ya existente.

Referencias bibliográficas

- [1] T. Mitchell, Machine Learning. 1146/annurev.cs, 1990.
- [2] API Standard 650, “Welded Tanks for Oil Storage”, Twelfth Edition March 2013, Addendum 1 September 2014, Errata 1 July 2013, Errata 2 December 2014, American Petroleum Institute, 2013.
- [3] ANSI/AWWA D100-11, “Welded Carbon Steel Tanks for Water Storage”, American Water Works Association, AWWA Standard, July 1, 2011.
- [4] New Zealand Society for Earthquake Engineering, “Seismic Design of Storage Tanks: 2009”, Recommendations of a NZSEE Study Group on Seismic Design of Storage Tanks, 2009.
- [5] Pineda P., Saragoni G. R. and Arze L. E. 2012. “Performance of Steel Tanks in Chile 2010 and 1985 Earthquakes”. Proceedings of the 7th International Conference on

- Behaviour of Steel Structures in Seismic Areas - STESSA, Santiago, Chile, pp. 337-342, February 9-11.
- [6] Pineda P. and G. R. Saragoni. 2015. "Backward Seismic Analysis of Steel Tanks". Proceedings of the 8th International Conference on Behaviour of Steel Structures in Seismic Areas - STESSA, Shanghai, China, pp. 337-342, July 1-3.
 - [7] Pineda P. and G. R. Saragoni. 2017. "Analysis of Steel Tanks in Chile Subduction Earthquakes". Proceedings of the 16th World Conference on Earthquake Engineering, 16WCEE 2017. Santiago, Chile, January 9th to 13th. Paper N°4452.
 - [8] Pineda, P & Saragoni G.R., "Diseño Sísmico de Estanques de Acero Basado en el Método Análisis Sísmico Backward". XII Congreso Chileno de Sismología en Ingeniería Sísmica - ACHISINA, Abril de 2019, Valdivia, Chile. Expositor
 - [9] Pineda, P. (2019) "Análisis Sísmico Backward de Estanques Atmosféricos de Acero". Tesis para optar al Grado de Magister en Ciencias de la Ingeniería. Mención Ingeniería Sísmica. Departamento de Ingeniería Civil. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile.
 - [10] Pineda, P. & Saragoni, G.R., "Tank Design Recommendations for Seismic Codes on Critical Industrial Facilities". 17th World Conference on Earthquake Engineering, 17WCEE 2020. Sendai, Japan, September 13th to 18th, 2020. Expositor.
 - [11] Pineda, P. & Saragoni, G.R., "Proposal of Seismic Coefficient and Estimate for Horizontal Sliding for Steel Tanks by Backward Seismic Analysis (BSA) Method". The 10th International Conference on the Behaviour of Steel Structures in Seismic Areas STESSA. Timisoara, Romania, 26-28 May 2021. Expositor.
 - [12] Anteproyecto Norma Técnica Chilena. Diseño Sísmico de Estanques Atmosféricos de Acero Apoyados en el Suelo. ICHA, Instituto Chileno del Acero. 2021 actualmente en desarrollo (presidente comité).
 - [13] P. Möller-Acuña, R. Ahumada-García and J. A. Reyes-Suárez, "Machine learning for prediction of frost episodes in the Maule region of Chile" in Ubiquitous Computing and Ambient Intelligence, Cham, Switzerland:Springer, pp. 715-720, 2017
 - [14] N. Japkowicz, Learning from Imbalanced Data Sets. Papers from AAAI Workshop, 21(9), 10–15, 1997. Retrieved from <http://www.aaai.org/Papers/Workshops/2000/WS-00-05/WS00-05-003.pdf>
 - [15]. Deshmukh, R. R. (2015). Data Cleaning: Current Approaches and Issues. (June).
 - [16] <https://cran.r-project.org>
 - [17] Hernández, J., Ramírez, M., & Ferri, C. (2007). Introducción a la Minería de Datos. España.
 - [18] Mchugh, M. L. (2013). The Chi-square test of independence Lessons in biostatistics. Biochemia Medica, 23(2), 143–149. Retrieved from <http://dx.doi.org/10.11613/BM.2013.018>
 - [19] Pedregosa, F., Weiss, R., & Brucher, M. (2011). Scikit-learn : Machine Learning in Python. 12, 2825–2830.
 - [20] Dagnino S., J. (2014). Análisis de varianza. Revista Chilena de Anestesia, 43(4), 306–310. Retrieved from <https://revistachilenadeanestesia.cl/PII/revchilanestv43n04.07.pdf>
 - [21] I RishAn empirical study of the naive Bayes classifier. JCAI Workshop on Empirical Methods in AI., 1997
 - [22] <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
 - [23] L. Breiman Random forests. Mach Learn. 45:5–32, 2001.

- [24] <https://cran.r-project.org/web/packages/RWeka/RWeka.pdf>
- [25] Wang QQ, Yu SC, Qi X, Hu YH, Zheng WJ, Shi JX, Yao HY. [Overview of logistic regression model analysis and application]. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2019 Sep 6;53(9):955-960. Chinese. doi: 10.3760/cma.j.issn.0253-9624.2019.09.018.
- [26] Xuan P, Sun C, Zhang T, Ye Y, Shen T, Dong Y. Gradient Boosting Decision Tree-Based Method for Predicting Interactions Between Target Genes and Drugs. *Front Genet*. 2019 May 31;10:459. doi: 10.3389/fgene.2019.00459. PMID: 31214240; PMCID: PMC6555260.