

Aplicación de inteligencia artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

Ogosi Auqui José Antonio¹, Lira Camargo Jorge², Guadalupe Mori Víctor Hugo³, Lira Camargo Zoila Rosa⁴, Cacsire Castillo María Pilar⁵

c18793@utp.edu.pe; jlira@unfv.edu.pe; victor.guadalupe@upsjb.edu.pe;
zlira@unab.edu.pe; maria.cacsire@upsjb.edu.pe

¹ Universidad Tecnológica del Perú, 07001, Lima Perú

² Universidad Nacional Federico Villareal, 07001, Lima Perú

³ Universidad Privada San Juan Bautista, 07001, Lima Perú

⁴ Universidad Nacional de Barranca, 07001, Lima Perú

⁵ Universidad Privada San Juan Bautista, 07001, Lima Perú

Pages: 152-160

Resumen: La gestión de aguas residuales es una cuestión de gran importancia para la sociedad actual. Dicha importancia, la cual se ha visto incrementada en los últimos años por la escasez de recursos hídricos, se puede observar además desde diferentes puntos de vista, el económico, el ambiental, el sanitario, etc. Actores fundamentales en la gestión de este recurso son las empresas proveedoras del servicio. Así, el que éstas tengan un buen desempeño se vuelve un aspecto clave para garantizar la calidad del servicio y, para que puedan lograr el citado buen desempeño, el que cuenten con salud financiera resulta clave. En este sentido, disponer de una herramienta que permita prever la bancarrota de este tipo de empresas con años de antelación resultaría de gran utilidad para evitar a los stakeholders, administración pública, entorno.

Palabras-clave: Estrés financiero, empresas gestoras de servicios de aguas residuales, inteligencia artificial.

Application of Artificial Intelligence for the prediction of financial stress in companies in the wastewater collection and treatment sector

Abstract: Wastewater management is an issue of great importance for today's society. This importance, which has increased in recent years due to the scarcity of water resources, can also be observed from different points of view: economic, environmental, sanitary, etc. Fundamental actors in the management of this resource are the companies that provide the service. Thus, their good performance becomes a key aspect in guaranteeing the quality of the service and, in order for them to achieve this good performance, it is essential for them to be financially healthy.

In this sense, having a tool that would make it possible to foresee the bankruptcy of this type of company years in advance would be very useful to prevent stakeholders, public administration, the environment.

Keywords: Financial stress, wastewater service management companies, artificial intelligence.

1. Introducción

El tratamiento de aguas residuales es un eslabón importante en el ciclo del agua que permite el saneamiento y la reutilización de ésta, facilita la generación de energía y permite la recuperación de los productos de los desechos (Morales, Reyes y Montealegre, 2015). En este sentido, la gestión y capacidad de adaptación de los recursos hídricos se ha convertido en una necesidad estratégica de carácter mundial, donde la importancia que han adquirido las aguas residuales ha llegado a tal magnitud que la edición correspondiente al año 2017 del Informe Mundial de las Naciones Unidas sobre el Desarrollo de los Recursos Hídricos se ha centrado exclusivamente en este recurso tan valioso para la sostenibilidad del medio ambiente (ONU, 2017).

Sin embargo, no todos los países tienen capacidad para gestionar correctamente las aguas residuales, su tratamiento está relacionado con el nivel de ingresos del país, en promedio, los países con ingresos altos tratan cerca del 70% de las aguas residuales que generan, mientras que esa proporción cae al 38% en los países de ingresos medios-altos y al 28% en los países de ingresos medios-bajos (Martín, 2017). En países de bajos ingresos, solo el 8% de las aguas residuales industriales y municipales se someten a algún tipo de tratamiento (González, 2017).

Es evidente que la gestión de aguas residuales es costosa y requiere de una importante inversión en infraestructura, por lo que la falta de recursos financieros constituye una limitación importante en el desarrollo de este sector y, por tanto, en la sostenibilidad del medio ambiente (Belda, 2019). De ahí que en la mayoría de los países industrializados fuesen las empresas públicas las encargadas de llevar a cabo esta actividad por el claro beneficio social que producían (Diez, 2021). No obstante, poco a poco se han ido introduciendo empresas privadas o se han ido privatizando empresas públicas con la intención de mejorar la eficiencia de dichas empresas (Tsai & Yen, 2014).

Por tanto, desde el ámbito privado, se hace necesario comprobar la viabilidad del sector de aguas residuales, de manera que no sólo sea interesante desde el punto de vista social, sino empresarial (Du Jardin, 2017). La necesidad de conocer si las empresas de este sector son financieramente autosuficientes o el posible deterioro financiero o bancarrota que se pueda producir por diversas circunstancias, permite, por un lado, alentar la inversión privada, mientras que por otro lado supone un desafío para los organismos públicos cuyo objetivo debería centrarse en desarrollar normativas para la promoción del sector (Armstrong, 2014).

Por otra parte, la bancarrota es un evento crítico que puede causar grandes pérdidas a los distintos stakeholders, por ello, el desarrollo de modelos efectivos de predicción es una tarea importante, a la vez que complicada (Bauer & Agarwal, 2014). El objetivo de los modelos de predicción de la bancarrota es predecir con cierta antelación si una empresa irá a la bancarrota o no (Hsu, Chang & Lin, 2016). Si el modelo predictivo no

presenta un buen rendimiento, podría provocar la toma incorrecta de decisiones tanto en la propia empresa como en los stakeholders y, como consecuencia de ello, una grave crisis (Chen, 2013). Similar al objetivo de la predicción de la bancarrota, ocurre con el estrés financiero, es decir, un buen modelo predictivo para el estrés financiero puede ayudar a tomar decisiones financieras con menor riesgo para los stakeholders (Divsalar & Roodsaz, 2012).

Por lo tanto, el presente trabajo tiene como objetivo mejorar el conocimiento de un sector de actividad económica con unas características especiales ya que, si bien en los últimos años, se ha desarrollado mucha literatura entorno a este tipo de empresas, estos se han centrado principalmente en aspectos técnicos como el empleo de nuevas tecnologías, procesos y materiales y ninguno en predecir la futura quiebra o estrés financiero (Stankovic et al., 2020). Este tipo de predicciones ya se hace en otras empresas como las manufactureras. Otro de los objetivos que se plantean es realizar una comparativa entre tres modelos con distinto número de variables predictivas para comparar cuál de ellos presenta resultados más exactos (Nouri & Soltani, 2016).

Por otra parte, la aplicación de la inteligencia artificial en las empresas suele implicar ideas novedosas. Puede ayudar a automatizar procesos y tareas que antes se realizaban de forma rutinaria y manual. Con la ayuda de máquinas inteligentes, estas tareas se pueden programar y completar mucho más rápido con menos recursos.

Además, el aumento de la asistencia tecnológica permite a los trabajadores del sector de tratamiento y recolección de aguas residuales realizar un trabajo de mayor calidad. Esto se debe a que la IA les permite acceder a más información, lo que les ayuda a concentrarse en su trabajo. Este mayor enfoque les permite tomar mejores decisiones cuando trabajan.

Asimismo, esta información permite a las empresas de la misma categoría tomar decisiones de manera más rápida y eficiente. Al proporcionar datos estructurados, esto permite generalizar los resultados.

SVMs utilizan un modelo lineal para implementar límites de clases no lineales a través de algunos vectores de entrada de mapeo no lineales en un espacio de entidad de alta dimensionalidad. El modelo lineal construido en el nuevo espacio puede representar un límite de decisión no lineal en el espacio original. En el nuevo espacio, se construye un hiperplano de separación óptimo. Así, SVM se conoce como el algoritmo que encuentra un tipo especial de modelo lineal, el hiperplano de margen máximo (Pisner & Schnyer, 2020). El hiperplano de margen máximo da la máxima separación entre las clases de decisión. Los ejemplos de entrenamiento más cercanos al hiperplano de margen máximo se llaman vectores de soporte. Todos los demás ejemplos de entrenamiento son irrelevantes para definir los límites de la clase binaria. SVM es lo suficientemente simple como para ser analizado matemáticamente, ya que se puede demostrar que corresponde a un método lineal en un espacio de características de alta dimensional no linealmente relacionado con el espacio de entrada. En este sentido, SVM puede servir como una alternativa prometedora que combina las fortalezas de los métodos estadísticos convencionales que son más teóricos y fáciles de analizar, y más métodos de aprendizaje automático, libres de distribución y robustos (Zendehboudi, Baseer & Saidur, 2018).

En este estudio, la función kernel utilizada fue la “función radial basis”. Además, hay dos parámetros asociados a la “función radial basis” que juegan un papel muy importante en el desarrollo de la ejecución del SVM: C (cost) y γ (gamma). C es el parámetro de penalización del término de error. Y γ (gamma) es el coeficiente de la función kernel, que en este caso es la “función radial basis”. La incorrecta selección de estos parámetros puede causar problemas de “overfitting” o “underfitting” (Okwuashi & Ndehedehe, 2020).

La teoría del conjunto difuso proporciona un modelo matemático para emular el proceso cognitivo humano. En comparación con los modelos probabilísticos o estadísticos, la teoría del conjunto difuso busca clasificar el razonamiento subjetivo y asignar grados de posibilidades para llegar a conclusiones (Jayalakshmi et al., 2021). Proporciona un medio para tomar decisiones basadas en datos ambiguos, imprecisos o incompletos. FL representa modelos o conocimiento usando reglas “if-then” en forma de “si A y B, entonces C”. Las cualidades de este método incluyen su simplicidad, rendimiento satisfactorio y robustez. Sin embargo, la construcción de modelos difusos de grandes y complejos sistemas es una tarea difícil que exige la identificación de muchos parámetros y el éxito va a depender de la elección de estos parámetros y de la función de membresía o pertenencia (membership function). Una forma de resolver este problema es utilizar un método de optimización, en este caso, se ha utilizado la técnica de los AG (Arji et al., 2019).

Este sistema neuronal ha demostrado su eficacia en dominios reales, incluyendo clustering, reconocimiento de patrones, reducción de dimensiones y la extracción de características. El modelo SOM es una red neuronal no supervisada (unsupervised) y se compone de dos capas neuronales. La capa de entrada tiene tantas neuronas como variables, y su función es simplemente capturar la información. La capa de salida tiene forma rectangular con “x” filas e “y” columnas (dos dimensiones). Cada neurona en la capa de entrada está conectada a cada neurona en la capa de salida a través de los pesos sinápticos (Zhu et al., 2020). Por lo tanto, cada neurona en la capa de salida tiene n conexiones a la capa de entrada. Cada una de estas conexiones tiene un peso sináptico asociado con ella. SOM intenta proyectar el espacio de entrada multidimensional, que en nuestro caso podría ser información financiera, en el espacio de salida de tal manera que los patrones de entrada cuyas variables presenten valores similares aparezcan cercanos entre sí en el mapa que se crea. Cada neurona aprende a reconocer un tipo específico de patrón de entrada. Las neuronas que están cerca en el mapa reconocerán patrones de entrada similares cuyas imágenes, por lo tanto, aparecerán cerca unas de otras en el mapa creado. De esta manera, la topología esencial del espacio de entrada se conserva en el espacio de salida. De este modo, es posible establecer en un mapa bidimensional diferentes zonas tales empresas activas y no activas y crear trayectorias de quiebra (Clark, Sisson & Sharma, 2020).

A la combinación de múltiples clasificadores se le llama “classifier ensembles”. El concepto de “ensemble” es una de las últimas tendencias en las líneas de investigación asociada a las técnicas de clasificación y tiene como objetivo la obtención de clasificadores de mayor exactitud combinando técnicas menos precisas. Básicamente se proponen para mejorar el rendimiento de clasificación de un único clasificador, ya que la combinación compensa los errores cometidos por clasificadores individuales en diferentes partes del espacio de entrada (Alzubi et al., 2020). El concepto es muy sencillo, un número determinado de

clasificadores comparten los datos de entrada (inputs) y sus respectivos datos de salida (“outputs”) son combinados (media aritmética, voting, bagging o boosting) para crear un único “output”. De esta forma, unas técnicas se complementan con otras y se obtienen mejores soluciones y más robusta (Tama, Comuzzi & Rhee, 2019).

2. Metodología

A continuación, se detallan sucintamente las técnicas de IA utilizadas en el presente trabajo.

Las máquinas de vectores de soporte son algoritmos de aprendizaje supervisado que fueron utilizados para aplicaciones de clasificación, regresión y procesamiento de señales médicas. Su objetivo fue encontrar el mejor hiperplano de separación que separe los puntos de datos en dicho estudio. Esto está representado por los signos más y menos; cuanto más amplio sea el margen entre las dos clases, mejor. Las SVM usaron estos signos para determinar implícitamente qué separación era la mejor. Además, con el procesamiento del lenguaje natural y las aplicaciones de reconocimiento de imágenes y voz de SVM fueron posibles. El margen fue el ancho máximo de una región paralela a un hiperplano sin puntos de datos en él. Solo los problemas de separación lineal permitieron que el algoritmo encuentre dicho plano; de lo contrario, el algoritmo utiliza márgenes blandos al permitir que una pequeña cantidad de puntos de datos se clasifiquen incorrectamente.

El algoritmo fue utilizado para tratar datos cuantificados; sin embargo, también se pudo utilizar con información no precisa. Por ejemplo, el algoritmo procesó conjuntos que no están definidos por ‘estrés’ o ‘no estrés’. Esto fue posible porque el lenguaje natural que se estaba estudiando requería un método para procesarlo y usarlo. Por lo tanto, surgió la necesidad de utilizar el algoritmo en estudios futuros. Eventualmente, su propósito evolucionó hacia el uso de la lógica difusa para cuantificar los atributos contenidos en el lenguaje natural.

Dicho mapa auto organizado fue una herramienta usada para reducir los datos de alta dimensión a un mapa 2D. Esto permitió que las similitudes en los datos se vean más fácilmente, al igual que PCA. Sin embargo, los mapas auto organizados fueron diferentes porque presentaron estos mapas como mapas en lugar de dimensiones reducidas.

Este modelo compuesto utilizó una fusión ponderada de varios modelos de clasificación para hacer predicciones. En síntesis, la fusión de variados modelos clasificados el cual aumentó la predictiva capacidad para las predicciones del estudio realizado.

3. Resultados

Actual	Predicción	
	Estrés	No-estrés
Estrés	Verdaderos Positivos (VP)	Falsos Negativos (FN)
No-estrés	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Tabla 1 – Matriz de confusión

Para cada una de las cuatro técnicas y cada uno de los tres modelos. Al finalizar el proceso, se calcularon la sensibilidad, especificidad y exactitud con los datos acumulados en cada una de las matrices de confusión. La Tabla 2 muestra las medidas de rendimiento obtenidas con el fichero de validación (test)

Medidas Rendimiento	Prevalencia	Sensibilidad	Especificidad	Exactitud	Media Gcom.
Mod1-SVM	29,54	63,06	95,78	86,11	77,72
Mod1-CE	29,54	70,17	91,07	84,90	79,94
Mod1-FL	29,54	67,55	88,02	81,97	77,11
Mod1-SOM	29,54	89,54	98,71	96,00	94,01
Mod1-Promedio	29,54	72,58	93,48	87,25	82,33
Mod2-SVM	29,54	66,42	89,30	82,54	77,01
Mod2-CE	29,54	73,65	88,58	84,17	80,77
Mod2-FL	29,54	71,89	86,86	82,44	79,02
Mod2-SOM	29,54	97,77	98,31	98,15	98,04
Mod2-Promedio	29,54	77,43	90,76	86,83	83,83
Mod3-SVM	29,54	42,34	93,24	78,21	62,83
Mod3-CE	29,54	77,92	83,84	82,09	80,83
Mod3-FL	29,54	69,65	85,49	80,81	77,16
Mod3-SOM	29,54	93,01	96,78	95,67	94,88
Mod3-Promedio	29,54	70,73	89,84	84,20	79,71

Tabla 2 – Medidas de rendimiento de datos de validación

Se puede observar, el buen rendimiento de todas las técnicas (excepto la técnica SVM del modelo 3 (78,21%), el resto de las técnicas y modelos presentan una exactitud superior al 80%) y en particular, el excelente rendimiento de la técnica SOM, con una exactitud del 96,00%, 98,15% y 95,67% en los modelos 1, 2 y 3, respectivamente. Teniendo en cuenta que una de las técnicas que forman el CE en este estudio es el SVM, se podría presuponer que ésta arroja mejores resultados. Esto es así en los modelos 2 y 3 donde la técnica CE mejora el resultado de las técnicas individuales (84,17% vs 82,59% para el modelo 2 y 82,09% vs 78,21% para el modelo 3). Sin embargo, este hecho no sucede en el modelo 1, donde la exactitud de la técnica SVM es del 86,11% frente al 84,90% de la técnica CE. Para intentar explicar este resultado, se analizan los valores de la sensibilidad y especificidad, observándose que la primera es 4,71 puntos superior en la técnica SVM respecto de CE (95,78% vs 91,07%). Sin embargo, la sensibilidad es 7,11 puntos superior en la técnica CE (70,17% vs 63,06%), pero dado que la prevalencia del análisis es del 29,54, adquiere más peso la especificidad que la sensibilidad y esto hace que la exactitud sea mayor en el modelo que presenta mayor especificidad. Para confirmar esta situación, se observa que la media geométrica de la sensibilidad y especificidad es superior en la técnica CE que en la de SVM.

Siguiendo con el análisis de sensibilidad y especificidad, se puede observar que en todos los casos en que la especificidad es superior a la sensibilidad, el valor mínimo de la especificidad es 83,84% (modelo 3, técnica CE), frente a un 42,34%, que es el valor mínimo de la sensibilidad (modelo 3, técnica SVM), por lo que se puede afirmar, que independientemente del modelo seleccionado, es un proceso más específico que sensible.

Si se observa los promedios de las medidas de rendimiento, se puede comprobar que la sensibilidad obtenida con modelo 2 es superior a la de los otros dos modelos (77,43%, 72,58% y 70,73% respectivamente). La especificidad del modelo 1 es superior al modelo 2 y 3 (93,40%, 90,76% y 89,84%). La exactitud del modelo 1 es superior al modelo 2 (87,25% vs 86,83%) y, por el contrario, la media geométrica del modelo 2 es superior al modelo 1 (83,83% vs 82,33%). Así pues, el modelo 1 es más específico y el modelo 2 es más sensible. Por otra parte, el modelo que peores medidas de rendimiento presenta es el modelo 3.

Por otra parte, el modelo 1 presenta mayor exactitud en las técnicas SVM y CE y, sin embargo, el modelo 2 es superior en las técnicas FL y SOM. En resumen, la mejor técnica predictiva es SOM en el modelo 2. Por otra parte, el modelo 1 va a realizar predicciones más exactas y específicas que el resto de los modelos y el modelo 2, es más sensible que el resto de los modelos. También se puede concluir que el modelo predictivo construido minimizando el error "BIC" (modelo 2) presenta mejor rendimiento que el modelo construido minimizando el error "AIC" (modelo 3).

La principal contribución de este estudio es la elaboración de tres modelos predictivos, dos de los cuales, modelo 1 y modelo 2, presentan un gran rendimiento para predecir el estrés financiero con dos años de antelación en estas empresas. También se puede observar que el modelo 2 es más sensible y presenta mayor media geométrica que el modelo 1 y, por otra parte, el modelo 1 presenta mayor especificidad y exactitud. La otra gran aportación está relacionada con las técnicas empleadas. Así, sin lugar a dudas, la técnica que mejores predicciones realiza es la SOM, es un modelo muy sensible y específico al mismo tiempo, seguido de CE. Por último, cabe destacar que el modelo obtenido minimizando el error "BIC" presenta mejor rendimiento que el obtenido minimizando el error "AIC", por lo que se puede afirmar que utilizando el error BIC para elaborar modelos predictivos, se obtienen mejores resultados que utilizando el error AIC.

4. Conclusiones

Como consecuencia de los buenos resultados obtenidos, se puede predecir con suficiente antelación las empresas que se encuentran en un riesgo elevado de entrar en bancarrota. Ello supone que tanto los stakeholders como los gestores de la propia empresa, cuentan con tiempo suficiente para tomar las medidas oportunas. Con esto se puede reducir la probabilidad de problemas posteriores de insolvencias con el consecuente abandono de la prestación del servicio, con las correspondientes consecuencias para las personas que reciben el servicio, las empresas que lo contratan (evitando tener que intervenir para garantizar el servicio), la propia empresa y hasta el propio medio ambiente. Por otra parte, aquellas empresas que no presenten riesgo financiero, seguirán teniendo la confianza de los acreedores y podrán obtener sin problemas la financiación para realizar

las inversiones necesarias en infraestructuras o innovación o tener la confianza de quien los contrate.

Referencias

- Alzubi, O. A., Alzubi, J. A., Alweshah, M., Qiqieh, I., Al-Shami, S., & Ramachandran, M. (2020). An optimal pruning algorithm of classifier ensembles: dynamic programming approach. *Neural Computing and Applications*, 32(20), 16091-16107.
- Arji, G., Ahmadi, H., Nilashi, M., Rashid, T. A., Ahmed, O. H., Aljojo, N., & Zainol, A. (2019). Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification. *Biocybernetics and biomedical engineering*, 39(4), 937-955.
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & physiological optics: the journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), 502-508. <https://doi.org/10.1111/opo.12131>
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432-442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Belda, I. (2019). *Inteligencia artificial*. RBA Libros.
- Chen, M.-Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180-195. <https://doi.org/10.1016/j.ins.2011.09.013>
- Clark, S., Sisson, S. A., & Sharma, A. (2020). Tools for enhancing the application of self-organizing maps in water resources research and engineering. *Advances in Water Resources*, 143, 103676.
- Diez Pérez, O. R. (2021). *Informe de vigilancia tecnológica: Inteligencia artificial*.
- Divsalar, M., & Roodsaz, H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, 523(March 2011), 504-523.
- Du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, 75, 25-43. <https://doi.org/10.1016/j.eswa.2017.01.016>
- González Martín, J. M. (2017). *Aplicación de inteligencia artificial en la detección de problemas financieros (Doctoral dissertation)*.
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*.
- Jayalakshmi, M., Garg, L., Maharajan, K., Jayakumar, K., Srinivasan, K., Bashir, A. K., & Ramesh, K. (2021). Fuzzy logic-based health monitoring system for covid'19 patients. *Cmc-Computers Materials & Continua*, 2430-2446.
- Martín, J. M. G. (2017). *Aplicación de inteligencia artificial en la detección de problemas financieros (Doctoral dissertation, Universidad de Las Palmas de Gran Canaria)*.

- Morales, I. R., Reyes, B. R., & Montealegre, J. G. (2015). Innovación tecnológica en el sector agropecuario (Doctoral dissertation, tesis de grado) Universidad Técnica de Machala).
- Nouri, B. A., & Soltani, M. (2016). Designing a Bankruptcy Prediction Model Based on Account, Market and Macroeconomic Variables, 9(1), 125-147
- ONU (2017). Informe Mundial sobre el Desarrollo de los Recursos Hídricos de las Naciones Unidas 2017: Las aguas residuales: el recurso desaprovechado, resumen ejecutivo; 2017.
- Okwuashi, O., & Ndehedehe, C. E. (2020). Deep support vector machine for hyperspectral image classification. *Pattern Recognition*, 103, 107298.
- Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Academic Press.
- Stankovic, M., Hasanbeigi, A., Neftenov, M. N., Ventures, T. I., Basani, M., Núñez, A., & Ortiz, R. (2020). Uso de tecnologías de la 4RI en agua y saneamiento en América Latina y el Caribe.
- Tama, B. A., Comuzzi, M., & Rhee, K. H. (2019). TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. *IEEE access*, 7, 94497-94507.
- Tsai, Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, 24, 977-984. <https://doi.org/10.1016/j.asoc.2014.08.047>
- Zendehboudi, A., Baseer, M. A., & Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production*, 199, 272-285.
- Zhu, G., Wu, X., Ge, J., Liu, F., Zhao, W., & Wu, C. (2020). Influence of mining activities on groundwater hydrochemistry and heavy metal migration using a self-organizing map (SOM). *Journal of Cleaner Production*, 257, 120664.