

Machine Learning para automatizar los sistemas de tickets de soporte: Una revisión literaria

Machine Learning to automate support ticket systems: A literature review

Recibido: octubre 08 de 2022 | Revisado: octubre 29 de 2022 | Aceptado: noviembre 10 de 2022

ALESSANDRO VENEGAS VILLARREAL¹
ESTELA VILLAR GARCÍA²
ALBERTO CARLOS MENDOZA DE LOS SANTOS²

RESUMEN

El sistema de tickets de soporte al día de hoy es un elemento clave para cualquier empresa. La clasificación de estos tickets puede ser una tarea compleja si se realiza manualmente, debido a que la asignación puede ser incorrecta y esto conducir a una reasignación de tickets, utilización innecesaria de recursos y extensión del tiempo de resolución. La tecnología Machine Learning aplicado en los sistemas de tickets de soporte ha logrado automatizar la clasificación de los tickets, esperando así una mejor asignación de los incidentes a resolver. Este artículo tiene como objetivo encontrar los algoritmos, basados en la tecnología Machine Learning, que obtengan una mayor tasa de precisión en la clasificación de tickets de soporte en la gestión de incidencias de tecnologías de información (ITIL).

Palabras clave: gestión de incidentes de TI, aprendizaje automático, tickets de soporte, ITIL, mesa de servicio

ABSTRACT

The support ticket system today is a key element for any company. Classifying these tickets can be a complex task if done manually, as the assignment can be incorrect, leading to re-assignment of tickets, unnecessary use of resources, and extension of resolution time. The Machine Learning technology applied to support ticket systems has managed to automate the classification of tickets, thus hoping for a better assignment of the incidents to be resolved. This article aims to find the algorithms, based on Machine Learning technology, that obtain a higher accuracy rate in the classification of support tickets in the management of information technology incidents (ITIL).

Keywords: IT incident management, Machine Learning, support tickets, ITIL, service desk

- 1 Universidad Nacional de Trujillo, Escuela de Ingeniería de Sistemas, Trujillo-Perú
2 Universidad Nacional de Trujillo, Escuela de Ingeniería de Sistemas, Trujillo-Perú

Autor para correspondencia E-mail:
evillar@unitru.edu.pe

© Los autores. Este artículo es publicado por la Revista Campus de la Facultad de Ingeniería y Arquitectura de la Universidad de San Martín de Porres. Este artículo se distribuye en los términos de la Licencia Creative Commons Atribución No-comercial – Compartir-Igual 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/>), que permite el uso no comercial, distribución y reproducción en cualquier medio siempre que la obra original sea debidamente citada. Para uso comercial contactar a: revistacampus@usmp.pe.

<https://doi.org/10.24265/campus.2022.v27n34.04>

Introducción

La evolución constante de las Tecnologías de la Información (TI) ha traído como consecuencia que las empresas necesiten adaptarse a dichos cambios, incluso si estas no están relacionadas directamente con las TI. Es aquí donde entra a tallar la gestión de servicios de TI (Information Technology Service Management - ITSM), la cual es una variedad de actividades dirigidas al mantenimiento de la infraestructura de TI (iZuev et al., 2018).

Por otro lado, la gestión de incidentes vendría a ser una de las diversas prácticas de la ITSM y es de suma importancia para poder facilitar y regularizar con prontitud el flujo de actividades de TI de una organización, ya sea que el soporte sea dado por la empresa misma o por un proveedor de TI, estas suelen implementar para sus usuarios, STS (Sistema de Tickets de Soporte), con el fin de que estos puedan crear tickets en los que describen su problema o solicitud (Simon et al., 2022).

El proceso de resolución que atraviesa un ticket vendría a estar definido por los siguientes tres pasos: (i) *Comprensión del ticket*: Este proceso hace referencia a la revisión y análisis de la información del ticket. (ii) *Asignación del ticket*: En este paso se da la clasificación y se establece el grupo de resolución. (iii) *Fijación del ticket*: En este paso se detalla la solución técnica para resolver el incidente reportado.

Estos pasos suelen ser elaborados por el personal a cargo del nivel 1 de la mesa de servicio, que usualmente son trabajadores temporales o menos calificados. Al ser este un proceso manual, además que, debido

a los esfuerzos de digitalización por parte de las empresas, el volumen de tickets entrantes suele ser grande, es que en muchas ocasiones pueden ocurrir errores, los cuales provocan principalmente un mal enrutamiento en los tickets, un aumento tanto en costes, como en tiempo de resolución del ticket y, por lo tanto, una disminución significativa en la satisfacción del usuario (Simon et al., 2022).

Es por este motivo que las empresas se encuentran realizando cada vez más esfuerzos por automatizar este proceso, de esta manera, cada vez es más común el uso de algoritmos de IA (Inteligencia Artificial) y ML (Machine Learning), dando así la posibilidad de automatizar tareas básicas, además de reducir los esfuerzos humanos y los errores del proceso (Qamili et al., 2018) apoyando o reemplazando a los miembros del personal de soporte de la mesa de servicio de primer nivel.

En su mayoría, las investigaciones previas se centran en propuestas de aplicación e implementación de algoritmos de ML en STS, sin embargo, son muy pocas las revisiones bibliográficas que proporcionan una descripción general del estado actual del presente campo de estudio. Dado este contexto, en la presente investigación se realizó una revisión de la literatura, con el fin de obtener un análisis de los algoritmos de ML más eficaces usados para la obtención de una mayor precisión en la clasificación y asignación de tickets de incidencias.

Método

Se llevó a cabo una revisión sistemática con base en la adaptación

de la metodología Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), ver (Urrútia & Bonfill, 2010). Aquí la pregunta de la investigación establecida para conducir el proceso metodológico fue la siguiente: ¿Qué algoritmos de Machine Learning dan mejores resultados en la clasificación de tickets de soporte en la gestión de incidentes?

Una revisión sistemática es un artículo de «síntesis de la evidencia disponible», en el que se realiza una revisión de aspectos cuantitativos y cualitativos de estudios primarios, con el objetivo de resumir la información existente respecto de un tema en particular (Manterola et al., 2013). Tomando en cuenta esta definición, se puede observar la importancia de desarrollar el estudio siguiendo un método repetible, entendible y que considere la evaluación del riesgo de sesgos.

Para dar inicio al proceso de búsqueda se emplearon ciertos términos que surgieron a partir de la pregunta de la investigación, tales como: “support ticket”, “incident management”, “itil”, “service desk” y “machine learning”. A continuación, se estableció la conexión entre los términos de búsqueda a través de operadores booleanos: (“support ticket” OR “incident management” OR “itil” OR “service desk” AND “machine learning”). Luego se determinó las bases de datos usadas como motor de búsqueda, las cuales fueron SCOPUS - (TITLE-ABS-KEY (support AND ticket) OR TITLE-ABS-KEY (it AND incident AND management) OR TITLE-ABS-

KEY (itil) OR TITLE-ABS-KEY (service AND desk) AND TITLE-ABS-KEY (machine AND learning) AND TITLE-ABS-KEY (algorithms)) AND PUBYEAR > 2017 AND PUBYEAR > 2017, SCIENCEDIRECT -(support ticket AND it incident AND service desk AND machine learning) Year:2018-2022e, y en ALICIA como “it incident” and “machine learning”IEEE XPLORE.

Criterios de inclusión y exclusión

En la presente investigación se adjuntaron artículos publicados en las bases de datos anteriormente mencionadas, únicamente en inglés y comprendidos entre los años 2018 y 2022.

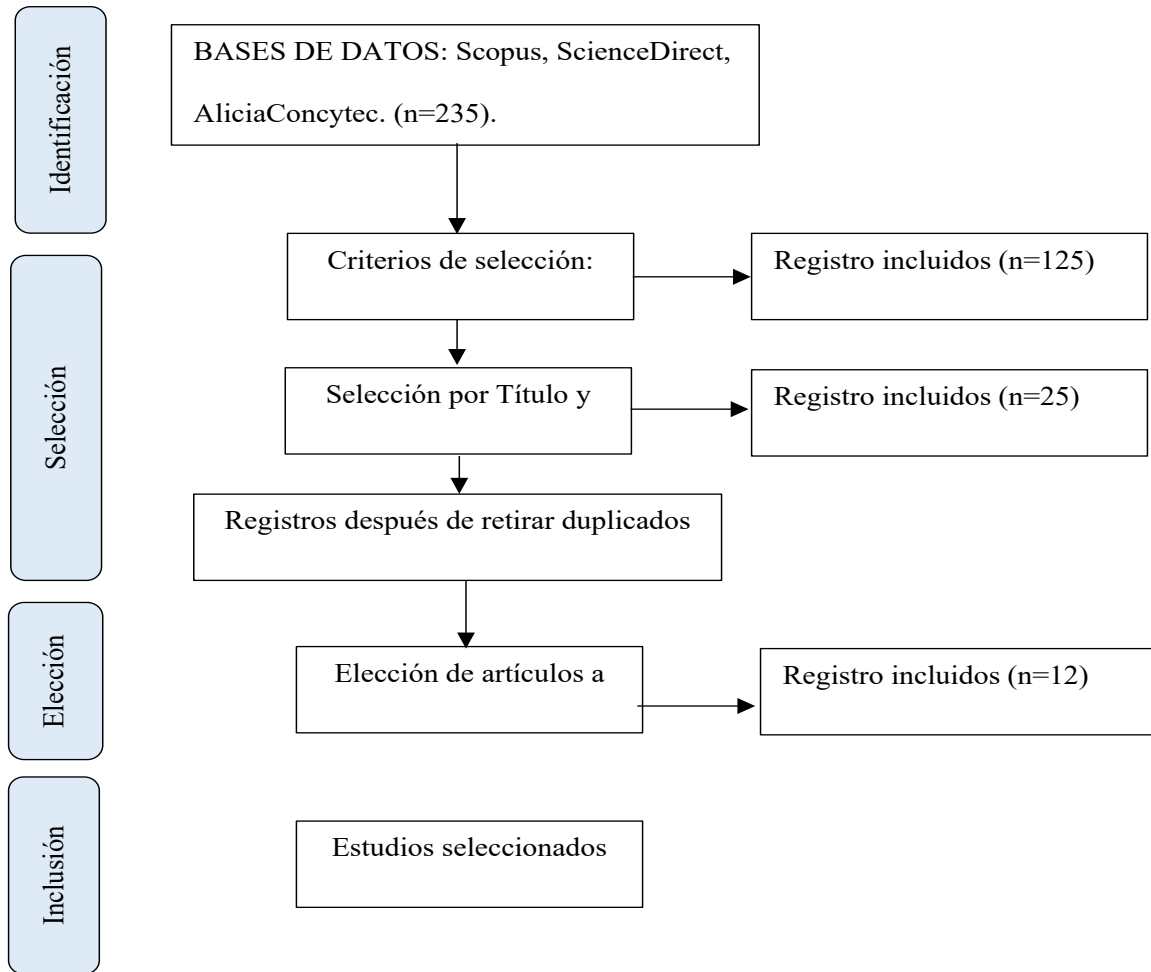
Criterio de inclusión

Se tomaron en cuenta los artículos que son aplicables en los entornos actuales de las tecnologías de la información, que se traten de investigaciones empíricas y no revisiones sistemáticas.

Criterio de exclusión

No se consideraron aquellas investigaciones sobre Machine Learning que no abarquen a los tickets de soporte o gestión de incidencias. También se descartaron aquellas investigaciones que apliquen Deep Learning en lugar de Machine Learning, que ahonden temas relacionados a la veterinaria, enfermería, artes y humanidades, economía, ciencias agrícolas y biológicas, química, ciencia medioambiental, física y astronomía, medicina, psicología o ciencias sociales.

Figura 1
Flujograma de discriminación de artículos encontrados



Resultados

De la búsqueda en primera instancia de los términos anteriormente mencionados en las bases de datos, se determinó un total de 235 artículos, ordenados de la siguiente manera: ScienceDirect con 39 artículos, Alicia con cuatro artículos y, por último, Scopus con 192 artículos. Sin embargo, a partir de este número total en la búsqueda, se aplicaron

los criterios de inclusión y exclusión mencionados anteriormente, obteniendo como resultado un total de 12 artículos escogidos.

En la Tabla 1, se muestran los 12 artículos incluidos en la revisión sistemática, estos se encuentran ordenados por base de datos y contienen los resultados, a modo de resumen, que cada uno de ellos obtuvo.

Tabla 1*Resúmenes de trabajos previos ordenados por Base de Datos*

Nº	Título	Base de datos	País	Resultado
1	Performance of 109 machine learning algorithms across five forecasting tasks: Employee behavior modeling, online communication, house pricing, it support and demand planning	Scopus	Bulgaria	Se realizó un análisis en diferentes dominios comerciales, para encontrar los mejores algoritmos en cada ámbito. Los algoritmos basados en Random forest funciona bien en un conjunto de datos de tamaño mediano.
2	Reducing Misclassification Due to Overlapping Classes in Text Classification via Stacking Classifiers on Different Feature Subsets	Scopus	Canadá	Se muestra como su método logra mejorar la tasa de precisión al momento de la clasificación.
3	Application of machine learning methods for automated classification and routing in ITIL	Scopus	Rusia	Este trabajo comparará diferentes métodos de aprendizaje automático para encontrar el enfoque más efectivo en la automatización del sistema de clasificación de tickets de soporte. Da como mejor a gradient boosting como técnica de Machine Learning que soluciona este problema.
4	Machine learning in incident categorization automation	Scopus	Portugal	Obtuvieron como resultado que los dos métodos con más alta tasa de precisión fueron: SVM y KNN.
5	Performance of Machine Learning Algorithms for IT Incident Management	Scopus	Indonesia	Este experimento usó la técnica de <i>validación cruzada</i> . Usaron distintos algoritmos de ML y obtuvieron como resultado que los mejores fueron: Random Forest, MLP y SVM.
6	Classifying the Unstructured IT Service Desk Tickets Using Ensemble of Classifiers	Scopus	India	Se realizó el experimento y obtuvo por resultado que los mejores algoritmos fueron: SVM y Random Forest.
7	IT Ticket Classification: The Simpler, the Better	Scopus	Hungría	En las pruebas realizadas, los algoritmos que obtuvieron mejores resultados en la puntuación F1, dados los rasgos lingüísticos a tomar en cuenta en los conjuntos de datos experimentales, fueron los <i>árboles de decisiones</i> y <i>naïve Bayes</i> , con un 100% en ambos casos.
8	An Intelligent Framework for Issue Ticketing System Based on Machine Learning	Scopus	Suiza	El algoritmo <i>SVM (Support Vector Machine)</i> arrojó un 75% en la puntuación F1. Esa puntuación mide la cantidad y calidad en la precisión de igual manera.
9	Machine Learning-based Automated Problem Categorization in a Helpdesk Ticketing Application	Scopus	Indonesia	El algoritmo que arrojó mayor porcentaje de precisión (82%) al clasificar un ticket aplicando el algoritmo <i>random forest</i> .

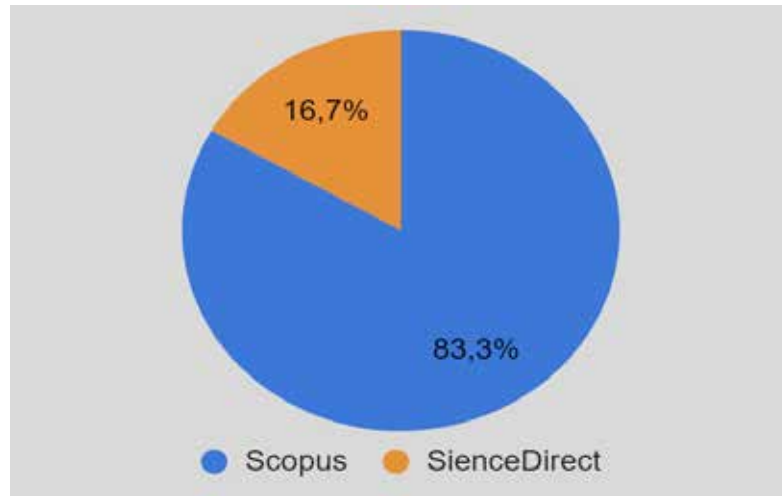
10	Automated IT service desk systems using machine learning techniques	Scopus	India	Los algoritmos usados en esta investigación fueron: Support Vector Machine (SVM), Regresión Logística (LR), Multinomial Naive Bayer (MNB) y kNN, obteniendo como resultados en el porcentaje de precisión 87%, 81%, 69% y 67%, respectivamente. Es decir, el algoritmo con mayor puntaje de precisión fue SVM.
11	Machine Learning in IT Service Management	ScienceDirect	China y Rusia	Se trabajó con diferentes métodos como Naive Bayesian clasificador, modelo de árboles de decisión de regresión logística y aumento de gradiente. Como respuesta a sus experimentos, el modelo Gradient Boosting Decision Trees funciona mucho mejor que otros.
12	A machine learning based help desk system for IT service management	ScienceDirect	Arabia Saudita	Se realizaron tres experimentos. Como resultado se obtuvo una mayor precisión con SMO.

En la Tabla 2, después del análisis de los estudios previos se pudo obtener los algoritmos recurrentes agrupados por autores.

Tabla 2

Resumen de evaluación

Nº	Autores	SVM	Logistic Regresion	Naive Bayes	Random Forest	D-Tree	kNN
1	(Gerunov, 2022)				✓		
2	(Wahba et al., 2022)		✓				
3	(Nikulín et al., 2021)						
4	(iZuev et al., 2018)	✓		✓		✓	
5	(Al-Hawari & Barham, 2021)	✓		✓		✓	
6	(Silva et al., 2018)	✓					✓
7	(Prihandono et al., 2020)	✓			✓		
8	(Paramesh et al., 2018)				✓		
9	(Revina et al., 2020)			✓		✓	
10	(Qamili et al., 2018)	✓	✓	✓	✓		
11	(Miliano et al., 2020)			✓	✓		
12	(Paramesh & Shreedhara, 2019)	✓	✓			✓	✓

Figura 2*Porcentaje de estudios incluidos por Base de Datos para la revisión sistemática*

En la Figura 2 se puede visualizar las Bases de Datos y los respectivos porcentajes de artículos que fueron aportados para esta revisión: Scopus cuenta con el mayor porcentaje de artículos seleccionados, con 83,3% y ScienceDirect con un 16,7%, todos en inglés.

Machine Learning es una rama de la inteligencia artificial que es usada, ampliamente, para la clasificación de datos, partiendo de un modelo de entrenamiento con el fin de obtener la predicción de nuevos resultados de datos (Prihandono et al., 2020). Algunos ejemplos de algoritmos de Machine learning son: Random Forest, SVM, Perceptrón Multicapa, Árboles de decisión, Regresión logística, etc. *Tickets de soporte*, es el registro de alguna observación a un servicio realizada por el cliente, de esta manera dándose posible organizar y encontrar cualquier demanda en el menor tiempo y de la manera más sencilla.

Un ticket de soporte podría estar contenido por un título, una descripción

general e información especial, que suele llamarse metadatos, tales como la prioridad del ticket, categoría del incidente, una identificación del ticket, etc. (Simon et al., 2022), en ocasiones también suelen tener comentarios (Al-Hawari & Barham, 2021). Por otro lado, un sistema de tickets de soporte vendría a ser una herramienta tecnológica que nos permitiría centralizar en un solo lugar (base de datos) la información de los tickets abiertos por los clientes. *Gestión de incidentes de TI*, ITIL la gestión de incidentes podría definirse como el proceso de responder a ciertos eventos inesperados o interrupciones al servicio, con el fin de restaurar el flujo normal del servicio, minimizando el impacto comercial. Para que este proceso pueda realizarse con éxito, es necesario cumplir con cuatro factores críticos de éxito: resolución rápida de incidentes, el mantenimiento de la calidad del servicio de TI, una mejora en la productividad de TI y del negocio y, por último, la constancia en la satisfacción del usuario (Silva et al., 2018).

Discusión

La automatización del sistema de tickets de soporte basado en Machine Learning se desarrolla e implementa desde el 2018, cada vez teniendo más presencia. De la Tabla 2, los algoritmos más usados en los artículos revisados fueron SVM (4, 5, 6, 7, 10, 12), Árboles de decisiones (4, 5, 9, 12), Naive Bayes (4, 5, 9, 10, 11), Random Forest (1, 7, 8, 10, 11), Regresión logística (10, 12) y kNN (1, 6, 12), ya sea que estos estén dados en su forma base o en alguna adaptación según el modelo a emplear.

No es posible concluir o generalizar diciendo que alguno, por ejemplo, SVM es el más *efectivo de todos*, debido a que, según la revisión realizada, durante el preprocesamiento del texto, que es un paso previo a la clasificación o categorización del ticket de soporte, las investigaciones previas presentan casos de uso particulares, enfoques y ámbitos distintos, etc.

Además, los modelos elaborados en las investigaciones estudiadas no apuntaban a establecerse como un clasificador universal, es decir, para cualquier conjunto de datos, esto debido a que dichos modelos siempre tenían que ajustar determinados parámetros para poder lograr una mejor optimización en la predicción según el caso de uso específico que estaban tratando. Sin embargo, los algoritmos pueden resultar más beneficiosos en ámbitos específicos, como, por ejemplo, la investigación (Qamili et al., 2018) obtuvo que algoritmos basados en SVM también eran óptimos para la detección de spam y el análisis de sentimientos.

Otro ejemplo, en la investigación (Gerunov, 2022), nos dice que los algoritmos basados en Random Forest (RF) parecen funcionar muy bien en conjuntos medianos de datos en donde hay diferentes predictores, la cual es una situación común. Por ejemplo, la investigación (Gerunov, 2022), nos menciona que mientras un conjunto de datos sea más escaso, se beneficiaría más con un modelado algorítmico cada vez más sofisticado, que pueda compensar los datos faltantes.

Es importante mencionar que la correcta definición de las categorías de incidencias, tiene una fuerte repercusión en la asignación de tickets mediante el algoritmo. Mientras mejor esté realizada la categorización de incidencias se puede obtener una alta tasa de precisión en la asignación de tickets. Es posible concluir que el uso de estos algoritmos aumenta las probabilidades de obtener resultados favorables en la clasificación automática.

Conclusiones

La clasificación tradicional de tickets de soporte es un problema en muchas empresas y aún más si existe un gran volumen de datos. En este estudio se pudo observar los distintos algoritmos de Machine Learning que se desarrollaron para facilitar la gestión a las mesas de servicio. Aplicando Machine Learning para la automatización de los Sistemas de Ticket de soporte, permite ganancias en la productividad, debido a que reducen la tasa de error, también ayuda a la rentabilidad del departamento de soporte.

Se encontró que cada modelo trabajado está diseñado para un caso

de uso particular en cual se tienen requisitos específicos y un tratamiento de la data tanto experimental como de entrenamiento acorde a los parámetros establecidos para ella. Los algoritmos más usados son: SVM, Random Forest, Árboles de decisión, Naive Bayes Regresión Logística y kNN. Se sugiere para estudios futuros que los criterios de inclusión y exclusión sean mejor estructurados, ello con el fin de realizar una obtención mejor filtrada

de los artículos acorde a los objetivos establecidos, evitando de esta manera la necesidad de descartar demasiados de forma manual. Además de ser también ser necesario e indispensable que se continúen realizando más investigaciones y actualizaciones sobre el uso del Machine Learning en la automatización de Sistemas de Tickets de Soporte, debido al aumento constante de data emitida en dichos tickets.

Referencias

- Al-Hawari, F., & Barham, H. (2021). A machine learning based help desk system for IT service management. *Journal of King Saud University - Computer and Information Sciences*, 33(6), 702-718. <https://doi.org/10.1016/j.jksuci.2019.04.001>
- Gerunov, A. A. (2022). Performance of 109 Machine Learning Algorithms across Five Forecasting Tasks: Employee Behavior Modeling, Online Communication, House Pricing, IT Support and Demand Planning. *Ikonicheski Izsledvania*, 31(2), 15-43. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85128755192&partnerID=40&md5=2d439dcd7fd3fc8c4b244ce2ea7543dc>
- iZuev, D., Kalistrato, A., & Zue, A. (2018). Machine Learning in IT Service Management. *Procedia Computer Science*, 145, 675-679. <https://doi.org/10.1016/j.procs.2018.11.063>
- Manterola, C., Astudillo, P., Arias, E., Claros, N. & Grupo MINCIR. (2013). Revisiones sistemáticas de la literatura. Qué se debe saber acerca de ellas. *Cirugía Española*, 91(3), 149-155. <https://doi.org/10.1016/j.ciresp.2011.07.009>
- Miliano, A., Steven, I., Kosim, K.P., Jayadi, R. & Mauritsius, T. (2020). Machine Learning-based Automated Problem Categorization in a Helpdesk Ticketing Application. *2020 8th International Conference on Orange Technology (ICOT)*, Daegu, Korea (South). <https://doi.org/10.1109/ICOT51877.2020.9468789>
- Nikulín, V. V., Shibaikin, S. D. & Vishnyakov, A. N. (2021). Application of machine learning methods for automated classification and routing in ITIL. *Journal of Physics: Conference Series*, 2091(1). <https://doi.org/10.1088/1742-6596/2091/1/012041>
- Paramesh, S. P., Ramya, C., & Shreedhara, K. S. (2018). Classifying the Unstructured IT Service Desk Tickets Using Ensemble of

- Classifiers. *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India. <https://doi.org/10.1109/CSITSS.2018.8768734>
- Paramesh, S.P., Shreedhara, K.S. (2019). Automated IT Service Desk Systems Using Machine Learning Techniques. In: Nagabhushan, P., Guru, D., Shekar, B., Kumar, Y. (eds) *Data Analytics and Learning. Lecture Notes in Networks and Systems*, (vol. 43, pp 331–346). Springer. https://doi.org/10.1007/978-981-13-2514-4_28
- Prihandono, M. A., Harwahu, R., & Sari, R. F. (2020). Performance of Machine Learning Algorithms for IT Incident Management. *2020 11th International Conference on Awareness Science and Technology (iCAST)*, Qingdao, China. <https://doi.org/10.1109/iCAST51195.2020.9319487>
- Qamili, R., Shabani, S. & Schneider, J. (2018). An Intelligent Framework for Issue Ticketing System Based on Machine Learning. *2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW)*, Stockholm, Sweden. <https://doi.org/10.1109/EDOCW.2018.00022>
- Revina, A., Buza, K. & Meister, V. G. (2020). IT Ticket Classification: The Simpler, the Better. *IEEE Access*, 8, 193380-193395. <https://doi.org/10.1109/ACCESS.2020.3032840>
- Silva, S., Pereira, R., & Ribeiro, R. (2018). Machine learning in incident categorization automation. *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, Caceres, Spain. <https://doi.org/10.23919/CISTI.2018.8399244>
- Simon, F., Clemens, D., & Holger, W. (2022). Improving Support Ticket Systems Using Machine Learning: A Literature Review. *Proceedings of the 55th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2022.238>
- Urrútia, G. & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica*, 135(11), 507-511. <https://doi.org/10.1016/j.medcli.2010.01.015>
- Wahba, Y., Madhavji, N. & Steinbacher, J. (2022). Reducing Misclassification Due to Overlapping Classes in Text Classification via Stacking Classifiers on Different Feature Subsets. In: Arai, K. (eds) *Advances in Information and Communication. FICC 2022. Lecture Notes in Networks and Systems*, (vol. 439, pp 406–419). Springer. https://doi.org/10.1007/978-3-030-98015-3_28