### Año de Publicación

De los 35 artículos seleccionados, el 71,4% de los documentos corresponden a artículos, mientras que el 28,6% restante corresponde a artículos de conferencia. Adicionalmente, se destaca que, durante el periodo de estudio de los últimos 10 años, las publicaciones pertinentes surgieron a partir del año 2015, como se visualiza en la figura 1.

El análisis revela un crecimiento en las publicaciones asociadas al tema de interés. En los primeros tres años, se observa un comportamiento constante con un total de 6 publicaciones. Sin embargo, a partir del 2018 y hasta 2019, no solo se experimenta un aumento en las publicaciones a 3 por año, sino que también se iguala al número de publicaciones de los tres primeros años en un periodo más corto. Adicionalmente, un incremento notable se evidencia a partir el 2019 y se extiende hasta el 2021 en donde se produce el pico más alto en publicaciones y aunque se aprecia una ligera disminución en 2022 y 2023 respecto a 2021, es importante mencionar que, hasta noviembre de 2023, se han registrado nuevas publicaciones relacionadas. Este patrón sugiere la posibilidad de un aumento continuo, lo que podría llevar a un repunte en las publicaciones para el año completo de 2023, superando, o al menos igualando, las cifras de 2022.

**Figura 1** *Número de publicaciones por año*

Finalmente, este fenómeno subraya la dinámica cambiante en la producción científica sobre el tema de interés. Este dinamismo se evidencia no solo en la variabilidad a lo largo de los 10 años de estudio, sino también en la tendencia más reciente en donde los últimos tres años han sido especialmente destacados. En este periodo, se ha publicado el 54,3% del total de documentos seleccionados.

***Países / Regiones***

La distribución geográfica de las publicaciones, como se observa en la figura 2, revela patrones interesantes en la contribución científica global. La región asiática lidera con un 57,69% del total de publicaciones en donde China sobresale con 12 contribuciones las cuales representan un 23.08% del conjunto total, seguida por India con 5 publicaciones, contribuyendo con un 9.62%. Otros países asiáticos como Bangladesh, Indonesia, Arabia Saudita, Corea del Sur y Taiwán también cuentan con una participación considerable, con 2 publicaciones cada uno, sumando un total de 9.23%. Además, países como Brunei Darussalam, Malaysia y Pakistán contribuyen con una publicación cada uno, representando un 5.77% adicional para el total de

**Figura 2**  
*Publicaciones por país*



Europa sigue en importancia, aportando un 28.85% del total de publicaciones. España y el Reino Unido lideran con 4 publicaciones cada uno, constituyendo un 7.69% respectivamente, mientras que Polonia contribuye con 3 publicaciones, representando un 5.77%. El resto de las publicaciones europeas se distribuyen entre Francia, Italia y Noruega, contribuyendo conjuntamente con un 7.69%.

África, aunque en menor proporción, presenta un interesante 7.69% del total de publicaciones, donde Kenia destaca con 2 contribuciones, y Mali y Nigeria aportan con una cada uno. Ahora, en cuanto a otras regiones, Australia, Chile y Canadá exhiben una publicación cada uno, representando en conjunto un 5,77% del total de contribuciones.

Finalmente, este análisis subraya la naturaleza global y colaborativa de la producción científica en torno al tema de interés en el cual los países asiáticos y europeos emergen como líderes destacados, liderando el frente de la investigación. Sin embargo, es notable la escasa representación latinoamericana, evidenciada por una única contribución proveniente de Chile.

### Áreas de Investigación

Es factible encontrar diversas áreas de conocimientos y aplicaciones, lo cual queda reflejado a través de un total de 16 áreas temáticas en las cuales cada investigación o artículo puede participar simultáneamente. En la tabla 1 se detallan algunas de las áreas temáticas relevantes para tener en cuenta en la ejecución del presente proyecto.

La mayoría de los documentos se clasifican principalmente en disciplinas como la ciencia de la computación, la ingeniería y las matemáticas. Estas disciplinas no solo brindan una base teórica y metodológica sólida, sino que son fundamentales para comprender y abordar la complejidad inherente de los conjuntos de datos desequilibrados. Este entendimiento se traduce en el desarrollo de modelos de aprendizaje automático ensamblados, los cuales ofrecen soluciones prácticas en diversos campos, abarcando desde la medicina hasta la toma de decisiones, la gestión de la energía, los negocios, la administración y contabilidad, así como en economía, econometría y finanzas.

La interconexión de estas disciplinas proporciona un enfoque integral que puede abordar problemáticas multidisciplinarias de manera eficiente.

**Tabla 1**  
*Publicaciones por área*

|  |  |  |
| --- | --- | --- |
| ***Área temática*** | ***Documentos*** | |
| Ciencias de la Computación | 29 |
| Ingeniería | 14 |
| Matemáticas | 8 |
| Medicina | 4 |
| Ciencias de la Decisión | 4 |
| Bioquímica, Genética y Biología Molecular | 3 |
| Ciencia de Materiales | 3 |
| Negocios, Administración y Contabilidad | 1 |
| Neurociencia | 1 |
| Economía, Econometría y Finanzas | 1 |
| Energía | 1 |

### Autores

En los 35 artículos seleccionados para el análisis bibliométrico, se identificaron 139 autores. Notablemente, el 98% de los autores contribuyeron con una única publicación en la muestra, mientras que el 2% de los autores se destacó al contribuir con dos documentos cada uno, incluyendo autores como Wang, Z., Shukla, S. y Ksieniewicz, P. En cuanto a citaciones, los artículos escritos por los autores Hu, Y.-H., Jhang, J.-S., Lin, W.-C. y Tsai, C.-F. han capturado considerable atención, siendo el artículo titulado "Clustering-based undersampling in class-imbalanced data" el más destacado, con un total de 479 citaciones hasta la fecha, seguido por "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy", con 186 citaciones, escrito por Galar, M., Herrera, F., Jeleń, L. y Krawczyk, B.

En la figura 3 se visualizan las relaciones y la fuerza total de los vínculos de co-autoría entre autores, calculadas mediante la herramienta VOSViewer, utilizando criterios tales con un mínimo de 9 citaciones y 1 documento por autor como mínimo. En este análisis, se identificaron 51 autores distribuidos en 14 clústeres. No se observa continuidad en la publicación de artículos entre los autores durante el periodo analizado. Sin embargo, se destaca que, entre los autores principales con más de una publicación con su respectiva cantidad de citaciones se encuentra: Wang, Z. (27) el cual es un autor contemporáneo con publicaciones en el 2022 al colaborar con Budhi, G.S. y Chiong, R, y aunque no se muestra en el gráfico, vale la pena señalar que continuó la producción en 2023 pero a la fecha aún no cuenta con citación alguna. Por otro lado, Shukla, S. (33) colaboró en el 2021 con Choudhary, R. y en el 2019 con Raghuwanshi, B.S. pero no alcanzó las citaciones mínimas para la visualización; mientras que Ksieniewicz, P. (9), solo realizó colaboraciones con Gulowaty, B.

**Figura 3**  
*Mapa de redes de coautoría presentado por VOSViewer.*

Diagrama

Descripción generada automáticamente

Adicionalmente, durante el 2020 existe el mayor número de colaboraciones, con un total de 8 autores participando en ellas, teniendo en cuenta que en promedio aproximadamente dichas colaboraciones son de 4 personas por artículo. Durante el periodo de 2015 a 2018, se observa consistentemente 2 colaboraciones por año, con excepción del 2017, que contó con solo 1 colaboración. Resaltando que, este último artículo ha sido el más citado hasta el momento.

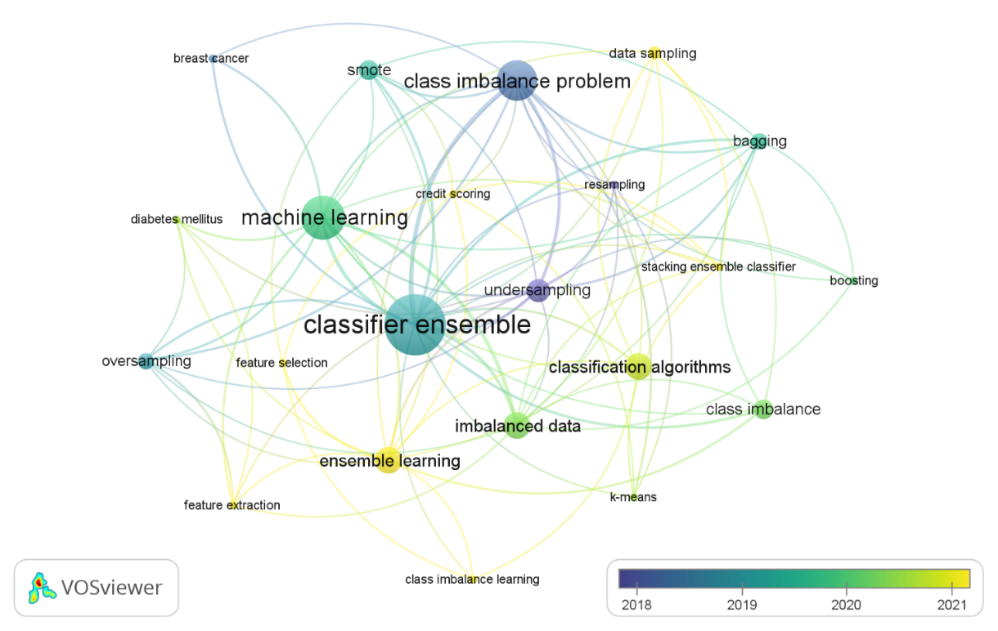
En contraste, se observa que en el 2019 no se presentó alguna colaboración en publicación de artículos que cumpliera con los criterios definidos para su visualización. Sin embargo, desde el 2020 hasta 2021, hubo un incremento en la cantidad de colaboraciones, pasando de 2 a 3, y finalmente en el 2022 se registraron 2 colaboraciones.

### Análisis de coocurrencia de palabras claves

Se identificaron un total de 85 palabras claves en los artículos seleccionados. De estas, se eligieron 23 palabras claves para su inclusión en la figura 4, siguiendo el criterio de que debían aparecer al menos dos veces en el conjunto total de artículos. Entre las palabras claves destacadas, se encuentra una alta frecuencia de términos como ‘*classifier ensemble*’, ‘*machine learning*’ y ‘*class imbalance problem’.* Es importante señalar que ‘*classifier ensemble*’ es el término más relevante y se encuentra estrechamente relacionado con otros conceptos, como ‘*bagging’*, ‘*boosting*’ o ‘*stacking ensemble classifier*’ que corresponden a métodos de aprendizaje automático ensamblado. Además, se encontraron términos como ‘*resampling*’, ‘*oversampling*’ y ‘*SMOTE*’, que están relacionados con técnicas de remuestreo de datos.

Durante el periodo analizado, el término *'classifier ensemble’* representó el concepto con mayor relevancia, destacando su importancia en el ámbito de la investigación, y durante los años 2020 y 2021 resaltan denominaciones relacionadas como *'ensemble learning' y ‘classification algorithms’.* Estos conceptos abarcan un marco conceptual que engloba tanto modelos de clasificación como los de regresión, evidenciado la amplitud de aplicaciones que abarca el aprendizaje ensamblado. Además, se identificaron técnicas adicionales, como *'feature extraction'*, *'feature selection'* y *'k-means'*, las cuales desempeñan un papel crucial en la fase de procesamiento de datos.

***Figura 4*** *Mapa de redes de coocurrencia por palabras clave presentado por VOSviewer*



En última instancia, se destaca que las aplicaciones más sobresalientes de este enfoque abarcan contextos diversos, como la evaluación crediticia y la identificación de enfermedades como la diabetes mellitus o el cáncer de mama. De esta manera, Este análisis de coocurrencia de palabras claves revela la amplitud y relevancia de los enfoques y las aplicaciones asociadas al aprendizaje ensamblado.