

# Primera Entrega Bases de Datos

Santiago Pulido Gómez  
Fundación Universitaria Konrad Lorenz

**Resumen**—Mediante google colab se realizó un análisis a los artículos sobre Machine-Learning hallados en la base de datos de scopus

## I. INTRODUCCIÓN

En la base de datos de scopus se desea realizar una investigación sobre Machine-Learning, para ello se hace uso de google colab en el cual se subió un código que nos permitiera realizar un análisis a los títulos, resúmenes y palabras clave de los artículos que le enviemos en un archivo .csv. Sin embargo es importante limitar esta cantidad a no más de 100 artículos para que dicho análisis no tarde tanto.

Por ello al momento de realizar la búsqueda sobre Machine-Learning en las bases de datos de Scopus, se emplea una serie de filtros con tal de disminuir el volumen de artículos para obtener un número óptimo y descargar los datos de estos dentro de un archivo .csv

## II. RESULTADOS

Luego de realizar los filtros en Scopus y haber almacenado los artículos obtenidos en el .csv, se procedió a subir dicho archivo a google colab para posteriormente ejecutar el código que analizará los títulos, los resúmenes y las palabras clave.

En el siguiente trabajo se tuvo en cuenta el análisis de 39 artículos sobre Machine-Learning, los cuales se analizaron. Generando una serie de gráficas que permiten hacer un análisis de forma visual de aquellas palabras que más se repiten en cada una de las secciones (títulos, resúmenes, palabras clave)

### II-A. Flujo del Analisis

Flujo del analisis bibliografico realizado.

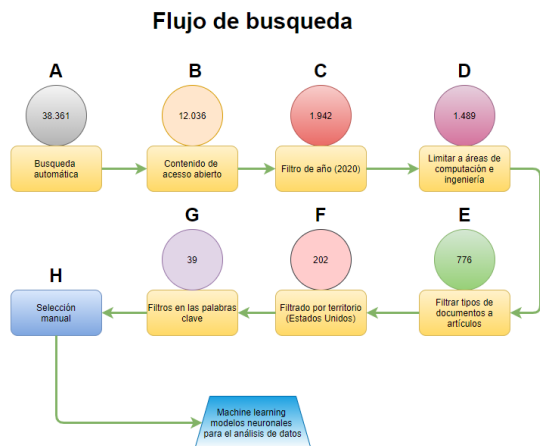


Figura 1. Flujo del filtro de los artículos.

### II-B. Query Exploratoria

Explicación inicial de la query exploratoria

En la Query se filtraron los artículos de la siguiente manera: Primero mediante las palabras clave machine-learning y computacion, seguidas de limitar la búsqueda a contenido de acceso abierto publicado en el año 2020, cuyas áreas fueron computación e ingeniería. A su vez se limitó esta búsqueda exclusivamente a artículos en los Estados Unidos que contuvieran exactamente como palabras clave "Machine Learning" "Learning Systems", logrando filtrar un número significativo de 38,361 documentos a 39 documentos que concuerdan con la búsqueda realizada en la base de datos de Scopus

```

TITLE-ABS-KEY(machine-learning AND computing) AND (LIMIT TO (OA, "y")) AND (LIMIT TO (PUBYEAR, 2020)) AND (LIMIT TO (SUBAREA, "COMP") OR LIMIT TO (SUBAREA, "ENGI")) AND (LIMIT TO (DOCTYPE, "ar")) AND (LIMIT TO (AFFLCOUNTRY, "United States")) AND (LIMIT TO (EXACTKEYWORD, "Machine Learning")) AND (LIMIT TO (EXACTKEYWORD, "Learning Systems"))
  
```

Figura 2. Query final del filtro realizado.

### II-C. Analisis de Titulo

Tras realizar el análisis de los títulos, se evidencia que las palabras más usadas son: "learning", usando "machine", es decir, estos artículos están enfocados en el uso del aprendizaje de las máquinas para la realización de diversos trabajos. Igualmente, podemos observar palabras como: "datamodel", "neural", las cuales nos dan a entender que se utilizan modelos neuronales que servirán, para que la máquina vaya guardando similitudes que más adelante serán usadas en sus diversos trabajos o análisis de datos.

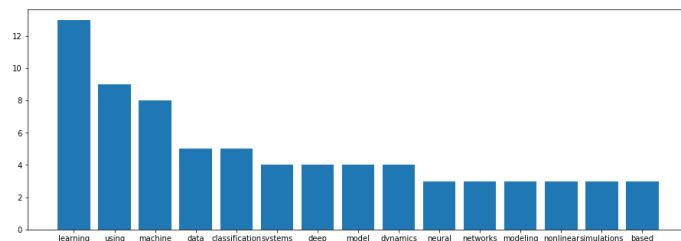


Figura 3. Gráfica títulos.

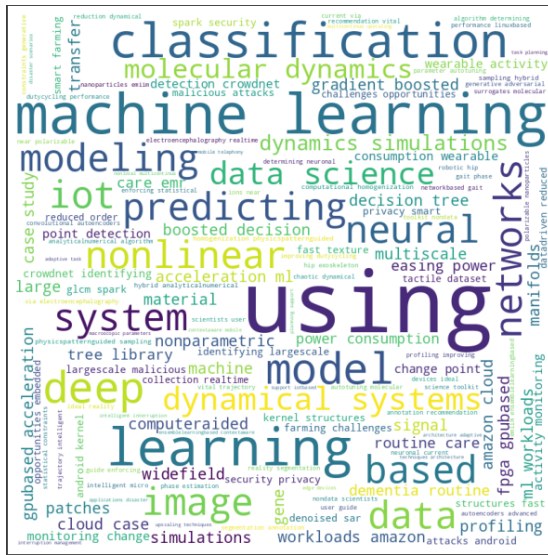


Figura 4. Imágen títulos más usados.

## II-D. Analisis de Keyword

Las palabras clave más utilizadas según el análisis realizado son: "learning", "machine", computing y "neural", las cuales permiten encontrar de manera más fácil los artículos que se utilizaron en este análisis. Permitiendo una mejor comprensión sobre el machine-learning que se intenta implementar en la computación desde diferentes ámbitos.

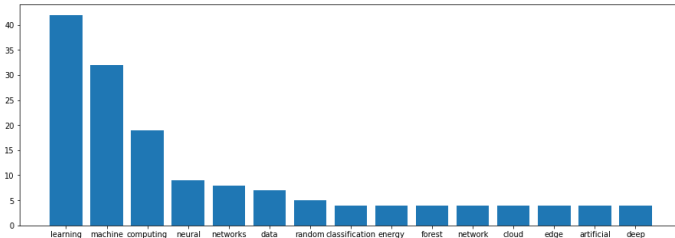


Figura 5. Gráfica palabras clave.

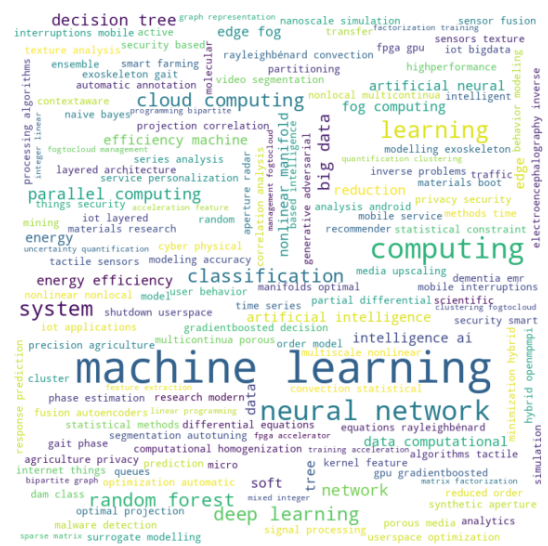


Figura 6. Imágen palabras clave más usadas.

## II-E. Analisis de Abstract

En los resúmenes de los artículos, las palabras más utilizadas son: "learning", "data", "model", "machine". Con lo cual nos podemos dar una idea de la importancia que tienen los datos con los que se alimentan el conocimiento de la máquina y el modelado de la red neuronal que utilizara la máquina para lograr retener dicha información. A partir de ella llevar a cabo las tareas que le hayan sido asignadas.

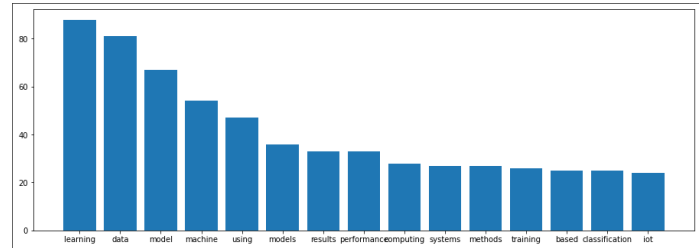


Figura 7. Gráfica palabras importantes en el resumen.

## II-F. Ranking de palabras clave para la construction del nuevo filtro

Posteriormente a los análisis, se puede concluir que las palabras más utilizadas en el ranking serian: "learning", "machine", "data", "model". Las cuales se pueden utilizar al momento de realizar una búsqueda en Scopus, para obtener un query más específico que nos permita filtrar de mejor manera los resultados.

## II-G. Query Resultante

El query resultante del análisis realizado, nos permite obtener desde un inicio tan solo 98 resultados.

TITLE-ABS KEY (machine-learning AND computing AND learning AND machine AND data AND model) AND (LIMIT-TO (OA, "AP")) AND (LIMIT-TO (AFRICOUNTRY, "United States")) AND (LIMIT-TO (PUBYEAR, 2022)) AND (LIMIT-TO (DOCTYPE, "AP")) AND (LIMIT-TO (SUBAREA, "COMP") OR LIMIT-TO (SUBAREA, "ENGR"))

Figura 8. Query final del análisis obtenido.

## III. CONCLUSIONES

1) Gracias al código que permite analizar de manera más y fácil efectiva los diferentes artículos entregados sobre un mismo tema, se logra obtener las palabras clave que nos permitirán reducir el número de artículos que nos aparecerán al momento de buscar en la base de datos de Scopus.

2) Las palabras clave obtenidas, permiten a la base de datos identificar de forma más clara y sencilla la búsqueda del tema del cual se desea indagar. Obteniendo de esta manera resultados más óptimos.

3) El uso de los diagramas permite visualizar de forma más sencilla los diferentes datos obtenidos, permitiendo comprarlos para saber cuáles se deben utilizar al momento de realizar la búsqueda de información.

## REFERENCIAS

Kleyko, D., Rachkovskij, D., Osipov, E., Rahimi, A. (2022). A Survey on Hyperdimensional Computing aka Vector Symbolic Architectures, Part I: Models and Data Transformations. ACM Computing Surveys, 55(6), 18-30.