



INSTITUTO POLITÉCNICO NACIONAL

**CENTRO DE INVESTIGACIÓN
EN COMPUTACIÓN**

Tarea 01

**Ideas iniciales, panorama histórico
y fuentes documentales confiables**

Edgar Fernando Espinosa Torres

Clasificación inteligente de patrones

Dr. Cornelio Yáñez Márquez



CIUDAD DE MÉXICO, 16 DE SEPTIEMBRE DE 2023.

Índice general

Introducción	1
Desarrollo y discusión	2
PARTE 1.	2
PARTE 2.	3
PARTE 3.	3
PARTE 4.	4
PARTE 5.	5
Conclusiones	6
Referencias bibliográficas	7

Introducción

Desde la primera revolución industrial, la investigación científica ha experimentado un acelerado impulso, generando una creciente necesidad de centralizar el conocimiento. Esta necesidad encontró una de sus soluciones con la llegada del internet, que propició la creación de plataformas diseñadas para reunir el extenso volumen de libros, escritos, artículos y ponencias que se generan constantemente. Estos sitios no solo agrupan la información, sino que también ofrecen herramientas para búsquedas rápidas y precisas. También se han creado repositorios de datos que tienen importantes usos en diferentes áreas como reconocimiento de patrones, machine learning, ciencia de datos, entre otros.

La accesibilidad que estas plataformas brindan a estudiantes y académicos permite poner la información a disposición de un público más amplio, lo que propicia debates más enriquecidos y fomenta el desarrollo de la ciencia al poder consultar en fuentes fidedignas.

En la presente tarea, se hará una breve revisión de algunas de las bases de datos bibliográficas más relevantes y repositorios de datos generales que pueden resultar de utilidad para el curso de clasificación inteligente de patrones.

Desarrollo y discusión

Propósito de la tarea:

Realizar una exploración inicial de los sitios web de búsqueda de documentos que nos proporciona el IPN; asimismo, dar un rápido vistazo a tres de los repositorios de datos que usaremos durante el curso CIP.

PARTE 1.

Ingresar a <https://www.scopus.com/>

1.a) Buscar un documento de interés personal (de cualquier tipo: libro, artículo, ponencia...) y reportar la ficha bibliográfica.

Rodríguez-Rodríguez I, Campo-Valera M, Rodríguez JV, Frisa-Rubio A. Constrained IoT-Based Machine Learning for Accurate Glycemia Forecasting in Type 1 Diabetes Patients. *Sensors* (Basel). 2023 Mar 31;23(7):3665. doi: 10.3390/s23073665. PMID: 37050725; PMCID: PMC10099355 [1].

1.b) Redactar un párrafo con tus primeras impresiones de las facilidades que brinda este importante sitio web.

Observé que SCOPUS proporciona una vasta cantidad de literatura académica de diversas disciplinas. Su interfaz es intuitiva y dispone de herramientas de búsqueda y filtros avanzados, permite que el usuario pueda localizar publicaciones y documentos de forma rápida. También informa sobre las métricas de citación, esto informa acerca del impacto y relevancia de las publicaciones.

PARTE 2.

Ingresar a <https://www.webofscience.com/wos/woscc/basic-search>

2.a) Redactar un párrafo con tus primeras impresiones de las facilidades que brinda este importante sitio web.

Me gustó que Web of Science permite rastrear citas de una manera sencilla, lo que permite que un investigador pueda identificar los trabajos más influyentes en su campo. Tiene varias funciones en su interfaz y métricas detalladas. Sin embargo, preferí la interfaz de SCOPUS por sentirla un poco mejor diseñada.

PARTE 3.

Ingresar a <http://archive.ics.uci.edu/ml/>

Responder brevemente las siguientes preguntas:

3.a) ¿Qué ofrece este repositorio?

Una importante colección de bases de datos y conjuntos de datos destinados a la investigación en *machine learning*.

3.b) ¿Cuál es la universidad que presta este servicio?

La Universidad de California, Irvine (UCI) [2].

3.c) ¿Quién fue el creador de este repositorio? ¿en qué año?

El repositorio fue creado por David Aha, quién era doctorante en el año de 1987 [2].

3.d) ¿Cuántos archivos de datos hay actualmente en el repositorio?

El repositorio contiene 644 archivos de datos.

3.e) ¿Cuál archivo de datos te llamó más la atención? Wine porque tiene como objetivo clasificar correctamente los vinos basándose únicamente en su composición química.

PARTE 4.

Ingresar a <http://www.keel.es/>

Responder brevemente las siguientes preguntas:

4.a) ¿Qué ofrece este repositorio?

Ofrece un software programado en Java que brinda una plataforma para la investigación en áreas de *data mining* y algoritmos evolutivos. Además, proporciona datasets diseñados de forma específica para distintas tareas que pueden ser utilizadas en *machine learning*.

4.b) ¿Cuál es la universidad que presta este servicio?

La Universidad de Granada, España. Cuenta con el apoyo del Ministerio de Educación y Ciencia de España [3].

4.c) ¿Quiénes fueron los creadores de este repositorio? ¿en qué año?

Fue desarrollado por investigadores del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Un personaje notable involucrado en su creación es el Dr. Francisco Herrera Triguero [3].

4.d) ¿Cuántos archivos de datos hay actualmente en el repositorio?

Contiene 592 archivos de datos.

4.e) ¿Cuál archivo de datos te llamó más la atención?

Ecoli porque trata de la clasificación de proteínas de la bacteria *Escherichia Coli* en función de la secuencia de aminoácidos.

PARTE 5.

Ingresar a <https://www.kaggle.com/>

Responder brevemente las siguientes preguntas:

3.a) ¿Qué ofrece este repositorio?

Ofrece competencias, datasets de una gran variedad de temas, cuenta con foros de discusiones y ofertas de trabajo ofrecidas por empresas externas.

3.b) ¿Quién fundó este repositorio y en qué año lo hizo? (busca en wikipedia)

Fue fundado por Anthony Goldbloom y Ben Hamner en el año 2010 [4].

3.c) ¿Cuántos archivos de datos hay actualmente en el repositorio?

Al día 10 de septiembre de 2023, enlista cerca de 250 mil archivos de datos disponibles.

3.d) ¿Cuál archivo de datos te llamó más la atención?

Me llamó la atención el archivo de datos Epicurious – Recipes with Rating and Nutrition porque no es común encontrar un *dataset* consistente en recetas de cocina.

PARTE 6. Redactar un párrafo con tu opinión acerca de la importancia que tendrán estos tres repositorios durante el curso CIP, al considerar las facilidades que brindan a estudiantes e investigadores.

Indudablemente contar con estos repositorios será de mucha relevancia para nuestro proceso de aprendizaje de la CIP ya que tenemos acceso a múltiples datasets y herramientas que permitirán que practiquemos e implementemos los algoritmos de clasificación. Incluso para los estudiantes que gustan de retos podrán hacerlo a través de Kaggle o participar en los foros de discusión, lo que puede ampliar y mejorar su experiencia en la materia.

Conclusiones

La era digital ha revolucionado la manera en la que los académicos y estudiantes pueden acceder, almacenar y compartir información. Las plataformas como SCOPUS y Web of Science atestiguan estas afirmaciones. Estos sitios son considerados de importante relevancia dentro de la comunidad científica pues permiten un acceso eficiente y bien estructurado a la literatura científica.

Por otro lado, dados los objetivos y la propia naturaleza del curso de Clasificación Inteligente de Patrones (CIP) es sumamente importante que los alumnos contemos con datasets robustos y provenientes de una amplia variedad de áreas. Los repositorios de UCI, KEEL y Kaggle nos brindaran las puertas para que podamos experimentar con sus datos, pero también documentarnos y ampliar nuestra experiencia, como en el caso de Kaggle, el cual cuenta con competencias y foros de discusión.

La combinación de acceso a literatura científica y conjuntos de datos diversos promueve en los estudiantes una formación sólida y orientada a la práctica en el campo de la CIP.

Referencias bibliográficas

- [1] Rodríguez-Rodríguez, I., Campo-Valera, M., Rodríguez, J. L. M., & Frisa-Rubio, A. (2023). Constrained IoT-Based machine learning for accurate glycemia forecasting in Type 1 diabetes patients. *Sensors*, 23(7), 3665. <https://doi.org/10.3390/s23073665>
- [2] *UCI Machine Learning Repository*. (s. f.). <https://archive.ics.uci.edu/about>
- [3] Alcala-Fdez, J., Del Jesus, M. J., Garrell, J., & Sanchez, L. (2004). Proyecto KEEL: Desarrollo de una herramienta para el análisis e implementación de algoritmos de extracción. *ResearchGate*. <https://acortar.link/F9Xltj>
- [4] Wikipedia. (2023). Kaggle. *Wikipedia*. <https://en.wikipedia.org/wiki/Kaggle>