Manual de prácticas para la mejora de la enseñanza del aprendizaje máquina aplicado a la Ciencia de Datos a gran escala

Protocolo de titulación

Autora:
Mariana Michell Flores Monroy
Estudiante de la Licenciatura en Tecnologías para la Información en Ciencias
417063470
michell_monroy@comunidad.unam.mx

Asesor:
Dr. Sergio Rogelio Tinoco Martínez
Técnico Académico de la Licenciatura en Tecnologías para la Información en Ciencias
Escuela Nacional de Estudios Superiores Unidad Morelia
stinoco@enesmorelia.unam.mx

Coasesor:
Dr. Heberto Ferreira Medina
Tecnico Académico Titular B, TC, definitivo
Instituto de Investigaciones en Ecosistemas y Sustentabilidad, UNAM
hferreira@cieco.unam.mx

Morelia, Michoacán Abril 2021

Índice

1. Resumen

Este trabajo forma parte del proyecto PAPIME titulado "Propuesta de mejora a la enseñanza del aprendizaje automático aplicado a la Ciencia de Datos a gran escala" con el número de proyecto PE106021. Se propone la elaboración de un manual de prácticas para estudiantes del nivel licenciatura, que ayude a mejorar el conocimiento del Machine Learning (ML) y su aplicación en la ciencia de datos a gran escala (Big Data).

El proyecto estará basado en diseñar, construir e implementar una guía de prácticas dirigida a los estudiantes a partir de sexto semestre en adelante de la Licenciatura en Tecnologías para la Información en Ciencias o de otras licenciaturas de le ENES Morelia que cuenten con los conocimientos básicos del ML.

Para realizar dicho proyecto, se tomará en cuenta la opinión de alumnos y docentes de las diferentes licenciaturas dentro de la ENES Morelia, con respecto a cuales temas son los que consideran de mayor importancia y que se deben impartir dentro de las materias que utilizan el aprendizaje automático dentro del plan de estudios de la LTICS.

El fin de este proyecto es el de mejorar la formación académica de los estudiantes dentro de la LTICS así como mejorar la calidad de la enseñanza de este tema por parte de los docentes.

2. Justificación

En la actualidad existen diferentes métodos para el análisis de grandes volúmenes de datos, haciendo que las ciencias de la información tomen un papel relevante en nuestra sociedad. Debido a su importancia, dentro de la LTICs de la ENES Morelia existen materias orientadas a la ciencia de datos, en especial a los métodos del Machine Learning. Estas materias, al igual que las técnicas y herramientas que se utilizan en el ML, son de alta importancia para el estudiante ya que forman la base que se requiere para materias más complejas como redes neuronales.

Así también, actualmente en la LTICs, las asignaturas del área del ML se imparten de manera teórica y práctica, especialmente debido al cambio de paradigma motivado por la pandemia médica del SARS-CoV-2. No obstante lo anterior y debido a la complejidad de los temas abordados, el rendimiento estudiantil no es el ideal. Aunado a lo antes mencionado, la aplicación de los métodos del aprendizaje automático sobre grandes volúmenes de datos no es considerado dentro de los temarios de las diferentes asignaturas en el área. Por todo lo anterior, una práctica complementaria del alumnado, aplicada a problemas reales sobre el Big Data, reforzará el estudio y la comprensión de estos difíciles temas.

3. Hipótesis

En la Licenciatura en Tecnologías para la Información en Ciencias (ENES Morelia, UNAM) existe una cantidad considerable de estudiantes cuyo desempeño en las asignaturas del área del ML ha sido bajo debido a la complejidad de sus temas. Con el uso del manual de prácticas se espera que su desempeño mejore después de la intervención de este proyecto.

4. Objetivos

4.1. Objetivo General

Desarrollar un manual de prácticas para la enseñanza del Machine Learning, aplicado a gran escala, dirigido a estudiantes a partir de sexto semestre de la Licenciatura en Tecnologías para la Información en Ciencias o de otras licenciaturas de le ENES Morelia que cuenten con los conocimientos básicos del Machine Learning.

4.2. Objetivos Particulares

1. Desarrollar una encuesta para diagnosticar los conocimientos previos e intereses del alumnado.
2. Establecer los temas que se van a cubrir en el manual de prácticas, relacionados al ML y con fundamento en la encuesta aplicada.

3. Elaborar el marco teórico del manual de prácticas con base en los temas seleccionados.
4. Determinar los ejemplos prácticos del ML que se abordarán en el manual de prácticas, conciliando con los docentes de la LTICs de las asignaturas del área del ML, la pertinencia de las prácticas propuestas enfocadas al Big Data.
5. Implementar los ejemplos prácticos usando el lenguaje Python.
6. Realizar una prueba piloto del manual de prácticas.
7. Realizar el diagnóstico de los resultados de la intervención a través de una encuesta de salida.
8. Publicar los resultados en la página web del proyecto.

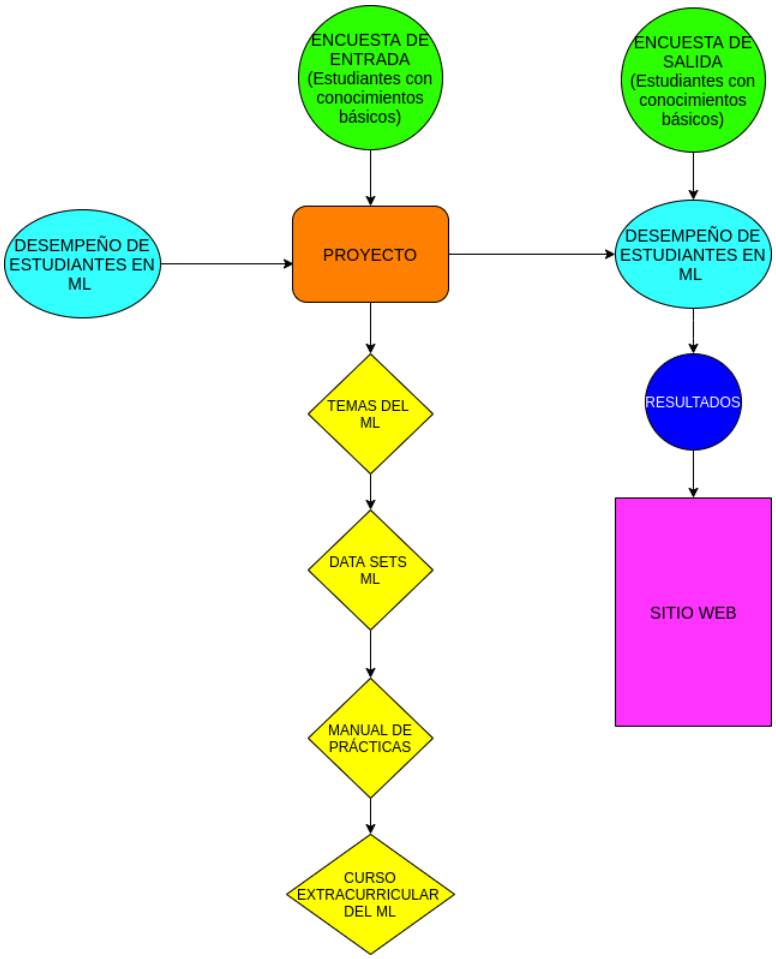

Figura 1: Mapa mental del desarrollo del proyecto.

5. Antecedentes

Según Kaplan (2016) se le llama Inteligencia Artificial (IA) a la ciencia que se encarga de crear sistemas inteligentes que sean capaces de imitar el comportamiento inteligente de los humanos para la toma de decisiones y lograr metas.

Haenlein y Kaplan (2019) describen que en el año 1942 se creía que la inteligencia artificial era un fenómeno futurista gracias al escritor Isaac Asimov y su obra *Runaround*.

Unos años después, el británico Alan Turing (1950) publicó su artículo *"Computing Machinery and Intelligence"* donde describe por primera vez cómo crear inteligencia artificial mediante las máquinas de Turing, además de explicar cómo se realiza la prueba de Turing para diferenciar IA de la inteligencia humana.

Mitchell (1997) establece que la IA se utiliza para resolver problemas complejos que la programación convencional no puede.

En su artículo, Jordan y Mitchell (2015), explican que dentro de la IA existe una rama llamada Machine Learning (ML) cuyo fin es mejorar el rendimiento de los algoritmos a través de su experiencia. Está técnica hace uso de herramientas como estadística, informática, matemáticas y ciencias computacionales y se fundamenta en el análisis de los datos.

La definición de Mitchell y cols. (1997) es la siguiente: "Se dice que un programa de computadora aprende de una experiencia E, con respecto a una tarea T y una medida de desempeño P, si su desempeño con respecto a T, medido por P, mejora con la experiencia E"

Lee y cols. (2017) hacen mención a Arthur Samuel, pionero en el estudio del ML, quien lo definió de la siguiente manera

"El aprendizaje automático es el campo de estudio que le da a las computadoras la habilidad de aprender, sin que esté explícitamente programada" Samuel (1959)

El término Machine Learning se acuñó oficialmente alrededor del año 1960, según lo relatado por Liu (2020). Este nombre consiste en la palabra "Machine", que hace alusión a cualquier dispositivo (robot, computadora) y la palabra "Learning" que hace referencia a la capacidad que se tiene de adquirir o descubrir patrones.

En la actualidad Mohri y cols. (2018) consideran al ML como la técnica de crear sistemas que sean capaces de aprender por sí mismos, utilizando grandes volúmenes de datos, haciendo que estos sean aptos para realizar análisis y, con ello, poder predecir futuros comportamientos.

5.1. Tipos de Machine Learning

El ML ha ganado importancia en las últimas décadas debido a su habilidad de realizar predicciones a partir de un conjunto de datos. Murdoch y cols. (2019) mencionan que los diferentes modelos de ML tienen la capacidad de adquirir conocimiento, relacionando características contenidas en los datos. A esto se le conoce comúnmente como "interpretaciones".

Géron (2019) explica que existen diferentes enfoques para el diseño de un sistema de ML. Estos se dividen en:

- Si están entrenados bajos supervisión humana o no. A estos se les denominan: **supervisado, no supervisado y de refuerzo.**
- Si pueden o no aprender sobre la marcha, denominados como **aprendizaje en línea.**
- Si detectan patrones de entrenamiento o si comparan nuevos datos con datos ya existentes. Este tipo de ML se cataloga como **aprendizaje basado en instancias o aprendizaje basado en modelos**

Aprendizaje Supervisado

El aprendizaje supervisado suele usarse cuando se cuentan con datos de los cuáles ya se sabe la respuesta que se desea predecir. Cunningham y cols. (2008) explican que este aprendizaje consiste en que el sistema pueda mapear entre los datos de entrada (input) y sus respectivas etiquetas (output) para después predecir las etiquetas dados nuevos datos no etiquetados, como se muestra en la Figura 2.

Según El Naqa y Murphy (2015) el principal objetivo es que el sistema aprenda a distinguir las características de una etiqueta de otra. El aprendizaje supervisado tiene la tarea de resolver los siguientes dos problemas Ayodele (2010), los cuales no pueden resolverse con programación simple:

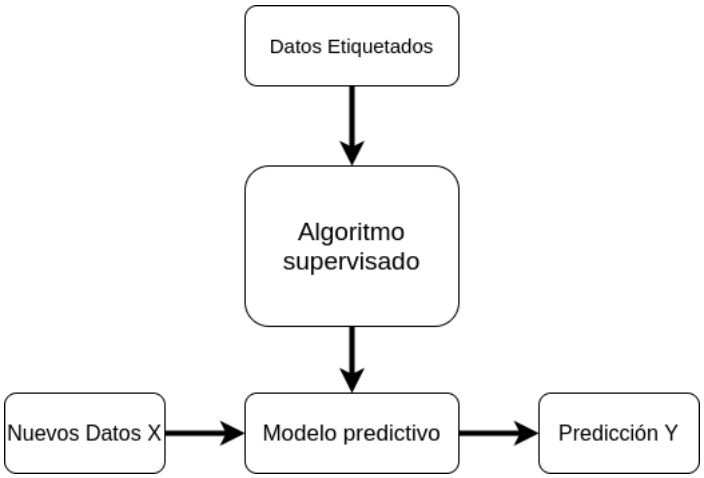

Figura 2: Diagrama del aprendizaje supervisado

- **Regresión.**
  Jagielski y cols. (2018) definen la regresión como un método en el cual se hacen uso de variables numéricas, para realizar predicciones, las cuales se espera que cada vez tengan menor margen de error. Estas variables se estudian para encontrar correlación entre ellas y sus respectivas etiquetas, para realizar predicciones de acuerdo a los patrones encontrados Montgomery y cols. (2021).
  Existen dos tipos principales de regresión: lineal y logística.

- **Clasificación.**
  La clasificación también se usa para hacer predicciones utilizando un conjunto de datos etiquetados, pero a diferencia de la regresión, la clasificación realiza predicciones discretas (llamadas clases) Li y cols. (2001).

  Kotsiantis y cols. (2007) mencionan los siguientes algoritmos que se usan para clasificación (entre otros):

  - Árboles de decisión.
    Este es un método muy utilizado debido a que es un algoritmo simple, fácil de comprender y porque no requiere de parámetros. Su y Zhang (2006) explican que el funcionamiento de los árboles de decisión consiste en un algoritmo

recursivo en el que en cada iteración escoge el atributo cuyo valor es más adecuado para dividir el conjunto de datos, hasta que todos los datos sean clasificados.

- k-vecinos más cercanos.
  Song y cols. (2007) explican que el algoritmo tiene la tarea de predecir la etiqueta de un dato (x_0) dados los k datos más cercanos, es decir, aquellos con menor distancia (euclidiana, distancia del coseno, etc.). Una vez que se tienen los k vecinos más cercanos, se revisan sus etiquetas y se le asigna a x_0 la etiqueta más repetida.
- Redes neuronales artificiales (RNA).
  Wang (2003) define a las RNA como un modelo que consiste en una capa de neuronas de entrada, algunas capas de neuronas ocultas y una capa de neuronas de salida. Cada conexión entre capas está asociada a un valor numérico (peso). También cuentan con funciones de activación, la más común es la función sigmoide.

Una de las aplicaciones de la clasificación, por ejemplo, es para detectar correo electrónico spam, como se ve en la Figura 3.

Figura 3: Un ejemplo de aprendizaje supervisado usado para clasificación de Spam

Aprendizaje No Supervisado

En el caso del aprendizaje no supervisado, los datos no están etiquetados, esto hace que el sistema tenga que aprender por sí mismo sin que se le indique si la clasificación es correcta o no Raschka y Mirjalili (2019)
El aprendizaje no supervisado, según Sathya y Abraham (2013), tiene la habilidad de aprender y organizar información, detectando patrones.

Para lograr clasificar los datos, se utiliza una técnica de agrupamiento o mejor conocida como clustering. Dayan y cols. (1999) proponen que el objetivo del clustering es agrupar datos cuyas características sean similares entre sí, como se ve en la Figura 4.

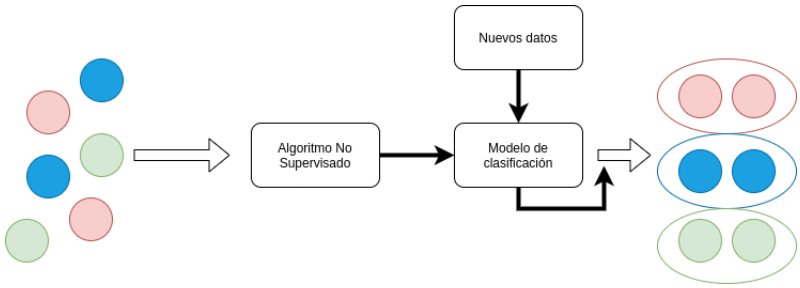

Figura 4: Un ejemplo de aprendizaje no supervisado usado para clustering, basado en los atributos de los datos.

Para implementar el clustering, Celebi y Aydin (2016) mencionan estas técnicas de aprendizaje no supervisado (entre otras):

- k-medias.
  Na y cols. (2010) explican que este algoritmo consiste en seleccionar aleatoriamente k centros, después calcular la distancia euclidiana (u otra métrica de distancia) de los demás datos para determinar cuál de los k centros es el más cercano y de esa forma clasificarlo en uno de los k grupos.
- Visualización y Reducción de Dimensiones.
  Vlachos y cols. (2002) mencionan que la reducción de dimensiones consiste en que un conjunto de datos reduzca su dimensionalidad sin la pérdida de información, para esto es común usar técnicas como Análisis de Componentes Principales (PCA en inglés) para que el conjunto de datos sea más fácil de procesar y de visualizar.

- Reglas de Asociación.
  De acuerdo con Aher y Lobo (2012), las reglas de asociación se usan comúnmente en minería de datos para encontrar de forma eficiente patrones o correlación en un gran conjunto de datos, para posteriormente obtener información de estos.

Aprendizaje Por Refuerzo
Wiering y Van Otterlo (2012) mencionan que el aprendizaje por refuerzo tiene como objetivo que el sistema aprenda en un entorno en el cual la única retroalimentación consiste en una recompensa escalar, la cual puede ser positiva o negativa (castigo).
La definición de Kaelbling y cols. (1996) es que el modelo recibe en cada iteración una recompensa r y el estado actual del entorno s, después el modelo toma una acción a de acuerdo con las entradas y eso es lo que se considera como la salida, la cual cambiará el estado s en la siguiente iteración.

En la Figura 5 se puede ver, de manera muy general, el comportamiento del aprendizaje por refuerzo.

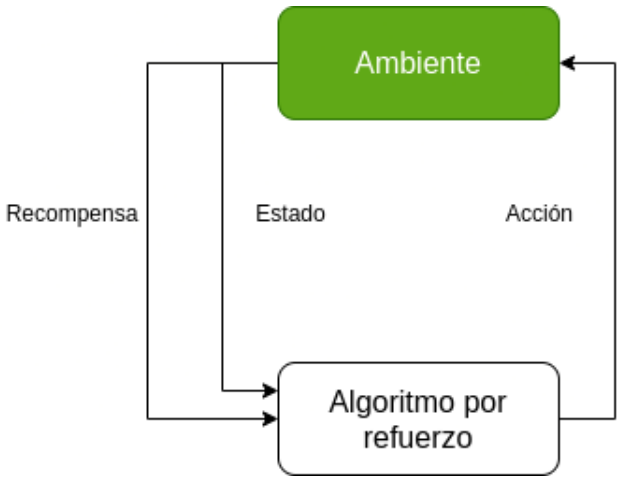

Figura 5: Diagrama de un algoritmo de aprendizaje por refuerzo.

En los últimos años, este algoritmo ha ganado terreno en el campo de investigación debido a sus aplicaciones mencionadas en Sutton y Barto (2018) tales como:

- Algoritmos que juegan ajedrez (Alpha Zero).

- Controlador adaptable de parámetros.
- Toma de decisiones.
- "Phil prepara su desayuno" el cual es un proceso de subtareas (como abrir el refrigerador, caminar a la estufa, romper un huevo, etc.) para lograr una tarea grande (preparar el desayuno).

5.2. Uso de ML en la actualidad

Una de las principales aplicaciones del ML en la actualidad es usar métodos supervisados para el análisis de grandes volúmenes de datos para obtener información sobre ellos. Por ejemplo en van Zoonen y Toni (2016) se muestra el caso de análisis de texto en redes sociales para entender la interacción entre usuarios.

Además de la utilidad del aprendizaje supervisado en el campo de la computación y matemáticas, como el desarrollo de Google DeepMinds y AlfaGo Siau y Wang (2018), existen otros usos en diferentes campos de la ciencia, por ejemplo:

- En Abbasi y Goldenholz (2019) se ve que debido a la gran utilidad de los algoritmos de ML para encontrar patrones, tienen un gran campo de aplicaciones tales como la detección de anomalías en electroencefalogramas.
- En biología, Libbrecht y Noble (2015) señalan que existen algoritmos encargados del análisis de grandes bases de datos de genomas, los cuales se entrenan para identificar potenciadores, nucleosomas, entre otras cosas.
- En medicina, Kourou y cols. (2015) indican que los algoritmos de ML se usan en la detección de tumores y su clasificación como benignos y malignos.
- Dentro de la psicología, Jiang y cols. (2020) los proponen como ayuda a la detección y predicción del riesgo de padecer algún trastorno mental.

6. Algoritmos de Machine Learning

En la sección 5.1 se habló sobre los tipos de algoritmos del ML. En esta sección se profundizará en los algoritmos que se usarán en

este manual de prácticas.

6.1. Árbol de decisión

Myles y cols. (2004) definen a un árbol de decisión como un algoritmo del tipo "divide y vencerás" usado comúnmente para hacer clasificación (aunque también puede usarse para regresión).
Su y Zhang (2006) proponen que el algoritmo inicia con un árbol vacío en el cual aún no hay nada de información acerca de los datos. Al ser un algoritmo avaricioso, este busca cual es el atributo que mejor divide el conjunto de datos y este se convierte en la raíz del árbol, después este proceso es recursivo, dividiéndose en subconjuntos que satisfacen la división de los datos.

A lo largo de los años, los investigadores han desarrollado diferentes algoritmos basados en árboles de decisión. Brijain y cols. (2014) explican que los modelos más importantes son los siguientes:

- **CHi-squared Automatic Interaction Detector (CHAID)**
  Este algoritmo fue desarrollado en 1980 por Gordon V Kass. Su objetivo es clasificar e identificar la interacción entre las variables con que trabaja. Una vez que detecta dicha interacción, CHAID selecciona el mejor atributo como nodo inicial, es decir, el atributo que mejor divide a los datos y realiza este procedimiento con todos los datos hasta que en los nodos finales los datos sean lo más homogéneos posibles entre sí.
  Agregado a eso, Rodríguez y cols. (2016) mencionan que CHAID es un proceso que no hace suposiciones sobre los datos. El algoritmo determina cuál es la mejor forma de combinar las variables para predecir un resultado binario. Esto lo hace dividiendo cada variable en subconjuntos mutuamente excluyentes basados en la homogeneidad de los datos.

  El criterio que se usa para determinar la división de los datos es *chi-squared test* (χ^2) Pandis (2016) explica que esta prueba solo muestra si existe una asociación entre variables, es decir, mide que tan dependiente es una variable de otra.
  La forma de calcular χ^2, mostrada en McHugh (2013) es la

siguiente:

$$\chi^2 = \sum_i^j \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

Donde:
O = Los valores observados.
E = Frecuencia esperada.

Para calcular la frecuencia esperada (E) se usa la siguiente formula:

$$E = \frac{M_R * M_C}{n} \tag{2}$$

Donde:
M_R = Suma de la fila.
M_C = Suma de la columna.
n = Número total de datos.
La ecuación 2 se aplica para cada uno de los datos y con el resultado de cada uno de ellos, se calcula χ^2 también para cada dato.

Como ejemplo tenemos la siguiente tabla 1

	Diestro	Zurdo
Hombre	24	13
Mujer	28	5

Cuadro 1: Tabla de datos

Lo primero que se tiene que hacer es la suma de las filas y las columnas, agregando una nueva fila y una nueva columna que contenga el total de las sumas. La tabla 2 quedaría de la siguiente forma:

	Diestro	Zurdo	Total
Hombre	24	13	37
Mujer	28	5	33
Total	52	18	70

Cuadro 2: Tabla con la suma de filas y columnas

Ahora se calcula la frecuencia esperada para cada valor en la tabla 2, usando la ecuación 2. Para el primer valor quedaría de

la siguiente forma:

$$\frac{52 * 37}{70} = 27{,}48$$

Esto se realiza para cada dato, quedando la tabla 3

	Diestro	Zurdo	Total
Hombre	27.49	9.51	37
Mujer	24.51	8.49	33
Total	52	18	70

Cuadro 3: Tabla con frecuencias esperadas

Ahora se aplica la ecuación 1 para calcular χ de cada dato. Para el primer dato en la tabla 1 quedaría de la siguiente forma:

$$\frac{(24 - 27{,}49)^2}{27{,}49} = 0{,}44$$

Se aplica el mismo proceso con cada uno de los valores obtenidos y la suma de ellos es la siguiente:

$$\chi^2 = 3{,}63$$

Este método ayuda mucho cuando se trata de análisis estadísticos.

- **Classification and regression tree (CART)**
  Los árboles de clasificación y regresión son descritos por Loh (2011) como métodos de ML que sirven para construir modelos capaces de realizar predicción de datos. Estos modelos se obtienen mediante la división recursiva de los datos y ajustado un modelo simple de predicción en cada una de esas divisiones.

  Este algoritmo usa una herramienta extra de aprendizaje, llamada "Poda". Loh (2014) explica que el método de poda es una herramienta bastante útil, ya que esta se basa en el concepto de eliminar al .[es]labón más débil". Estos eslabones están indexados por valores de costo-complejidad y eliminar los eslabones más débiles, se reduce el problema del sobreajuste, pero con un mayor costo de cálculo.

Estos valores de costo-complejidad se pueden medir usando los coeficientes de *Gini* y *Entropía*.

Gini: Este coeficiente es el más usado en árboles de clasificación, Alvaredo (2011) explica que este es más sensible a las transferencias en el centro de las distribuciones de datos.
? menciona que Gini utiliza la siguiete formla de impureza: i(t) = sum$_{kl}p(k|t)p(k|l)$

- **C4.5**

6.2. Modelos de Regresión

6.3. *k*-vecinos más cercanos (KNN)

6.4. Sistemas de recomendación

6.5. Spark

6.6. Sistema de archivos HDFS

7. Métricas

7.1. RMSE

7.2. MAE

Error absoluto medio (MAE por sus siglas en inglés) es una de las métricas más usadas para evaluar el desempeño de varios modelos de ML. Chai y Draxler (2014) menciona que MAE le da el mismo peso a todos los errores

$$\frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{3}$$

Donde:
y_i son las calificaciones reales.
$\hat{y}_i$ son las calificaciones predichas.
n el número de elementos en el conjunto de prueba.

Como se puede ver en la fórmula (**cambiar**) MAE mide qué tan cerca está la predicción en relación al valor real del conjunto de datos. Como mide el promedio de las distancias entre los valores reales

y las predicciones, un MAE "perfecto" es cuando el promedio de las distancias es 0. Es decir, las predicciones fueron iguales a los valores reales.

Una modificación que se tiene de MAE es NMAE, Jannach y cols. (2010) muestran que consiste en normalizar los valores del MAE con respecto a los valores con los que se trabaja.

$$\frac{MAE}{r_{max} - r_{min}} \tag{4}$$

Donde:
r_{max} es el valor máximo del conjunto de datos.
r_{min} es el valor mínimo del conjunto de datos.

7.3. Exactitud

7.4. Precisión

7.5. Recall

7.6. F1 score

7.7. Fβ

7.8. Sobreajuste

7.9. Subajuste

8. Formato de las Prácticas

9. Resultados

10. Discusión

11. Anexos

11.1. Prácticas

11.2. Datos

Los datasets que se usaron en las prácticas mostradas anteriormente, se pueden descargar de los siguientes links.

- Pasajeros del Titanic.
  https://www.kaggle.com/c/titanic/data
- California Housing.
  https://www.kaggle.com/fedesoriano/california-housing-prices-data-extra-features
- Pozos profundos del lago de Cuitzeo
  https://drive.google.com/file/d/19WVDYOM1xbF2hvbo75MDGe7OMN1clGhK/view?us
- Online Retail K-means & Hierarchical Clustering.
  https://www.kaggle.com/hellbuoy/online-retail-customer-clustering
- Anime Recommendation Database 2020.
  https://www.kaggle.com/hernan4444/anime-recommendation-database-2020
- RUOA de 2015 a la actualidad.
  https://ruoa.unam.mx/index.php?page=estaciones&id=9
- 100,000 UK Used Car Data set.
  https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes
- Información de datos públicos.
  Por decidirse

Referencias

Abbasi, B., y Goldenholz, D. M. (2019). Machine Learning Applications in Epilepsy. En (Vol. 60, pp. 2037–2047). Wiley Online Library.

Aher, S. B., y Lobo, L. (2012). A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning. *International Journal of Computer Applications*, *39*(1), 48–52.

Alvaredo, F. (2011). A note on the relationship between top income shares and the gini coefficient. *Economics Letters*, *110*(3), 274–277.

Ayodele, T. O. (2010). Types of Machine Learning Algorithms. *New Advances in Machine Learning*, *3*, 19–48.

Brijain, M., Patel, R., Kushik, M., y Rana, K. (2014). A Survey on Decision Tree Algorithm for Classification.

Celebi, M. E., y Aydin, K. (2016). *Unsupervised Learning Algorithms*. Springer.

Chai, T., y Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, *7*(3), 1247–1250.

Cunningham, P., Cord, M., y Delany, S. J. (2008). Supervised Learning. En *Machine Learning Techniques For Multimedia* (pp. 21–49). Springer.

Dayan, P., Sahani, M., y Deback, G. (1999). Unsupervised Learning. *The MIT Encyclopedia of The Cognitive Sciences*, 857–859.

El Naqa, I., y Murphy, M. J. (2015). What is Machine Learning? En *Machine Learning in Radiation Oncology* (pp. 3–11). Springer.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media.

Haenlein, M., y Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*(4), 5–14.

Jagielski, M., Oprea, A., Biggio, B., Liu, C., Nita-Rotaru, C., y Li, B. (2018). Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning. En *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 19–35).

Jannach, D., Zanker, M., Felfernig, A., y Friedrich, G. (2010). *Recommender systems: an introduction.* Cambridge University Press.

Jiang, T., Gradus, J. L., y Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687.

Jordan, M. I., y Mitchell, T. M. (2015). Machine Learning: Trends, Perspectives, and Prospects. *Science*, *349*(6245), 255–260.

Kaelbling, L. P., Littman, M. L., y Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Kaplan, J. (2016). *Artificial Intelligence: What Everyone Needs to Know.* Oxford University Press.

Kotsiantis, S. B., Zaharakis, I., y Pintelas, P. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3–24.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., y Fotiadis, D. I. (2015). Machine Learning Applications in Cancer Prognosis and Prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17.

Lee, A., Taylor, P., Kalpathy-Cramer, J., y Tufail, A. (2017). Machine Learning Has Arrived. *Ophthalmology*, *124*(12), 1726–1728.

Li, W., Han, J., y Pei, J. (2001). CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. En *Proceedings 2001 IEEE International Conference on Data Mining* (pp. 369–376).

Libbrecht, M. W., y Noble, W. S. (2015). Machine Learning Applications in Genetics and Genomics. *Nature Reviews Genetics*, *16*(6), 321–332.

Liu, Y. H. (2020). *Build Intelligent Systems Using Python, TensorFlow2, PyTorch and Scikit-learn*. Packt Publishing Ltd.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), 14–23.

Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329–348.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, *23*(2), 143–149.

Mitchell, T. (1997). Does Machine Learning Really Work? *AI Magazine*, *18*(3), 11–11.

Mitchell, T., y cols. (1997). Machine Learning. McGraw-hill New York.

Mohri, M., Rostamizadeh, A., y Talwalkar, A. (2018). *Foundations of Machine Learning*. MIT Press.

Montgomery, D. C., Peck, E. A., y Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., y Yu, B. (2019). Definitions, Methods, and Applications in Interpretable Machine Learning. *Proceedings of the National Academy of Sciences*, *116*(44), 22071–22080.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., y Brown, S. D. (2004). An Introduction to Decision Tree Modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275–285.

Na, S., Xumin, L., y Yong, G. (2010). Research on K-means Clustering Algorithm: An Improved K-means Clustering Algorithm. En *2010 Third International Symposium on Intelligent Information Technology and Security Informatics* (pp. 63–67).

Pandis, N. (2016). The Chi-square Test. *American journal of orthodontics and dentofacial orthopedics*, *150*(5), 898–899.

Raschka, S., y Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.

Rodríguez, A. H., Avilés-Jurado, F. X., Díaz, E., Schuetz, P., Trefler, S. I., Solé-Violán, J., ... others (2016). Procalcitonin (PCT) Levels for Ruling-out Bacterial Coinfection in ICU Patients with Influenza: a CHAID Decision-Tree Analysis. *Journal of infection*, *72*(2), 143–151.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229.

Sathya, R., y Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, *2*(2), 34–38.

Siau, K., y Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, *31*(2), 47–53.

Song, Y., Huang, J., Zhou, D., Zha, H., y Giles, C. L. (2007). Iknn: Informative k-Nearest Neighbor Pattern Classification. En *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 248–264).

Su, J., y Zhang, H. (2006). A Fast Decision Tree Learning Algorithm. En *AAAI* (Vol. 6, pp. 500–505).

Sutton, R. S., y Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433.

van Zoonen, W., y Toni, G. (2016). Social Media Research: The Application of Supervised Machine Learning in Organizational Communication Research. *Computers in Human Behavior*, *63*, 132–141.

Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G., y Koudas, N. (2002). Non-linear Dimensionality Reduction Techniques for Classification and Visualization. En *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 645–651).

Wang, S.-C. (2003). Artificial Neural Network. En *Interdisciplinary Computing in Java Programming* (pp. 81–100). Springer.

Wiering, M., y Van Otterlo, M. (2012). Reinforcement Learning. *Adaptation, Learning, and Optimization*, *12*(3).