

Evaluación de algoritmos de aprendizaje con datos públicos abiertos de machine learning mediante Orange3

Roxana Martínez, Pablo Vilaboa, Nelson Catala

Centro de Altos Estudios en Tecnología Informática (CAETI)
Universidad Abierta Interamericana (UAI)

Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina
{roxana.martinez, pablo.vilaboa}@uai.edu.ar
nelson.catala@alumnos.uai.edu.ar

Abstract. Existe una tendencia a nivel general por el impulso de la apertura de datos públicos por parte de los gobiernos. Esto conlleva a que no sólo es fundamental para el crecimiento de los países, sino que, además, incrementa la transparencia gubernamental para con los ciudadanos, y, por otra parte, es una forma de motivar a la utilización e implementación de la innovación tecnológica y a la participación ciudadana. El aporte de este trabajo de investigación conlleva a un relevamiento de los algoritmos de aprendizaje más relevantes en aspectos de aprendizaje supervisado como así también en un estudio general de las herramientas de machine learning más utilizadas en la actualidad. Como siguiente paso, este trabajo propone un análisis para la evaluación de algoritmos de aprendizaje de datos públicos abiertos, en este caso se toma en cuenta el estudio de un dataset público enfocado a enfermedades del corazón a nivel de salud. A través de la herramienta Orange se analizan los distintos algoritmos, y mediante una evaluación de testeo y puntuación (“Test and Score” y “Confusion Matrix”) se realiza la clasificación y ranqueo de los resultados arrojados para estos modelos según el criterio de mejor precisión en algoritmo.

Keywords: Machine Learning, Datos Públicos Abiertos, Aprendizaje Supervisado.

1 Introducción

1.1 Gobierno Abierto

Uno de los conceptos que se encuentra en pleno auge a nivel mundial es el de Gobierno Abierto, esto se debe a que existe un razonable consenso en la doctrina de que en esencia se refiere a una evolución del sistema democrático que se encuentra basada en el establecimiento de mecanismos para la transparencia de los gobiernos, así como de espacios permanentes de colaboración y participación de los ciudadanos [1]. En este aspecto, el rol de los datos públicos abiertos es fundamental, ya que la apertura de datos permite generar nuevas oportunidades en el que la información del sector público y su reutilización, se convierten en un instrumento generador de igualdad, brindan-

do nuevos derechos a los ciberciudadanos, por lo que esta disponibilidad de datos abiertos puede propiciar el desarrollo de factores de progreso en los ámbitos social, económico y cultural [2]. En estos tiempos, los datos son un recurso indispensable para cualquier actividad de gestión pública, por lo que es necesario mantenerlos actualizados, claros y completos [3].

1.2 Datos Abiertos y Machine Learning

La utilización de los datos abiertos permite el incremento de la interoperabilidad e integra a diversos sectores públicos, lo que conlleva a un aumento de la transparencia en la gestión pública de un país. Con el impulso de la innovación tecnológica, a partir de los datos, varios organismos tanto estatales como empresas privadas, generan iniciativas de valor agregado a través de los datos abiertos públicos disponibilizados en los portales de éstos.

En la misma línea entra en juego el enfoque de articulación de los datos abiertos con el concepto de machine learning, “este aprendizaje automático describe la capacidad de los sistemas para aprender de los datos de entrenamiento específicos del problema para automatizar el proceso de creación de modelos analíticos y resolver las tareas asociadas” [4]. El machine learning “es una rama dentro del campo de la Inteligencia Artificial que proporciona a los sistemas la capacidad de aprender y mejorar de manera automática, a partir de la experiencia” [5]. La obtención de esta información generada permite que un modelo realice predicciones de posibles escenarios. Es decir, la selección de algunos datos públicos permite comenzar con el proceso de aprendizaje que es implementado a través de un algoritmo, para analizar y luego, examinar los datos en búsqueda de patrones para culminar con un escenario final de cálculo de una determinada predicción.

Existen varios trabajos de investigación que se identifican en este contexto en varios ámbitos, algunos de ellos, por ejemplo, se centran en la salud como el caso de estudio de los autores Shailaja, K., Seetharamulu, B., & Jabbar, M. A., en el que el aprendizaje automático se utiliza para descubrir patrones a partir de fuentes de datos médicos y proporciona capacidades excelentes para predecir enfermedades [6]. Otro trabajo enfoca una propuesta llamada CORDATA [7], que es una aplicación web de gestión de datos abiertos para seleccionar inhibidores de corrosión. Otro trabajo de investigación [8] presenta el análisis de modelos predictivos de producción agroindustrial con Machine Learning a partir de fuentes de información pública, como otros autores que también presentan casos de estudio [9] con fuentes de datos públicos, explorando una metodología para el análisis de la violencia en el departamento de Bolívar mediante técnicas de machine learning, o el caso de aplicación de machine learning para análisis de los fenómenos de violencia intrafamiliar en el departamento del Atlántico [10] o bien trabajos relacionados con la evaluación de modelos de transporte mediante datos abiertos y técnicas de aprendizaje automático [11].

2 Estado del Arte

2.1 Categorías

Machine learning convierte los datos en información y facilita la toma de decisiones, su objetivo principal es identificar el objetivo del aprendizaje con sus correspondientes características de los datasets, para luego, seleccionar el tipo de aprendizaje que mejor se ajuste y pueda brindar una solución que responda a las necesidades.

En machine learning, existen tres grandes categorías [12]:

- **Enfoques de aprendizaje supervisado:** Estos algoritmos necesitan humanos para proporcionar la entrada y la salida requerida, además de proporcionar comentarios sobre la precisión de la predicción en el proceso de entrenamiento.
- **No supervisado:** Los enfoques de aprendizaje no supervisado Estos algoritmos no requieren de ningún proceso de capacitación. Pero, los enfoques de aprendizaje supervisado son más simples que los enfoques de aprendizaje no supervisado.
- **Semisupervisado:** Este enfoque utiliza datos de entrenamiento tanto etiquetados como no etiquetados.

Como se mencionó, el concepto de Machine learning se asocia regularmente al aprendizaje automático, esta estrategia radica en un conjunto de acciones que permiten, por medio del uso de algoritmos, identificar patrones. En este apartado analizaremos 6 algoritmos supervisados que fueron considerados para este estudio:

- **KNN:** El algoritmo de K vecinos más cercanos es un algoritmo sencillo, y por dicha sencillez es uno de los algoritmos más utilizados para la toma de decisiones, son simples y fáciles de aplicar y pueden ser utilizados para resolver problemas de clasificación y de regresión. Este algoritmo de tipo supervisado puede usarse tanto en valores discretos como continuos. Esencialmente, busca los valores similares para luego hacer conjeturas de nuevos valores [13]. Por otra parte, este algoritmo es basado en una instancia, lo que significa que no aprende estrictamente de un modelo, es decir, memoriza los valores en la etapa de entrenamiento para usarlas como base de conocimiento. Este conocimiento, es usado luego en la fase de predicción [14].
- **Tree:** Este algoritmo, puede utilizarse para problemas tanto de clasificación como regresión. Es un mecanismo que permite agrupar los datos, alimentando las etapas de aprendizaje y predicción. Como consecuencia de esto, el algoritmo de árbol de decisión permite predecir las características de un elemento observado o desconocido mediante la construcción de un modelo sostenido en los datos existentes [15]. Cabe mencionar que es necesario obtener un conjunto de datos para entrenar el modelo, por lo que, es posible representar un patrón para identificar un elemento nuevo que cumpla con dicho patrón [16].

- **Random Forest:** Surge como solución a una de las debilidades del algoritmo Tree, cuando estos tienen muchos niveles de profundidad. Esta estrategia consta de crear muchos árboles y que estos trabajen conjuntamente. El uso de este algoritmo implica el complemento de lógica de programación para crear el modelo y entrenarlo [15]. Por otra parte, se puede citar que, aunque parte del uso de un árbol de decisión, éste tiene un mejor rendimiento [17].
- **SVM:** Si consideramos un gran espacio de datos, este permite hacer una relación mutua o recíproca entre los datos del conjunto. Permite categorizar los datos observados e incluso si los datos no se pueden separar linealmente. Su estrategia, es encontrar un separador y con ellos distinguir cada una de las categorías [18]. Este algoritmo utiliza una función para la transformación conocida como función kernel, además, existen diferentes tipos que pueden enumerarse en lineal, polinómico, de base radial o sigmoide.
- **Naive Bayes:** Organiza un problema, mediante una cantidad y las relaciones que existen entre las variables observadas. Es decir, permite estimar la ocurrencia futura de variables no conocidas comparándolas con variables conocidas. Es una técnica que reduce la incertidumbre basándose en probabilidades. Esta práctica, construye un método basado en estadísticas con el fin de poder construir un modelo probabilístico [19]. Las redes Bayesianas, son métodos para obtener conocimiento, estas se basan en un modelo descriptivo, el cual se sostiene en estrategias centradas en el descubrimiento de relaciones de independencia o relevancia entre las variables, o bien, en un método predictivo, usando las redes como clasificadores. Esto último nos permite construir redes de muchos niveles mediante restricciones al proceso de aprendizaje [20].
- **Logistic Regresión:** Es un método estadístico, que modela una variable cualitativa de naturaleza binaria en función de algunas variables independientes. Crea un modelo de clasificación, y puede ser simple o múltiple. Al igual que otros que hemos comentado, es un método estadístico, y, además, se utiliza cuando el resultado solo pueden ser dos valores posibles [21]. Para terminar, se aplica la regresión logística cuando la regresión lineal no aporta solución. Esto se traduce en que, gracias a esta regresión, puede evaluar la probabilidad de que un evento ocurra o no.

2.2 Herramientas de Machine Learning

En la actualidad existen diversas herramientas de aprendizaje automático más populares y utilizadas. Estas herramientas usan diferentes lenguajes de programación y se ejecutan en ellos, algunos de estos se ejecutan en Python, en C++ o bien en Java.

A continuación, se muestran algunas de las más relevantes, siendo éstas:

- **Knime [22]:** es una herramienta de aprendizaje automático de código abierto que se basa en la GUI. No requiere ningún conocimiento de programación.

- **Accord.net [23]:** Accord.NET Framework es un marco de aprendizaje automático de .NET combinado con bibliotecas de procesamiento de audio e imágenes completamente escritas en C#. Es un marco completo para crear aplicaciones de visión por computadora, audición por computadora, procesamiento de señales y estadísticas de producción, incluso para uso comercial.
- **Scikit-Learn [24]:** Es un paquete de aprendizaje automático de código abierto, está construido sobre Python. Es una plataforma unificada, y de múltiples propósitos, con una comunidad de soporte muy participativa.
- **TensorFlow [25]:** Es una plataforma que va de extremo a extremo a código abierto para el aprendizaje automático. Esta plataforma tiene un ecosistema integral y flexible de herramientas, bibliotecas y recursos de la comunidad.
- **Weka [26]:** Se puede acceder a través de una interfaz gráfica de usuario. El software es muy fácil de usar. La aplicación de esta herramienta es en investigación y docencia y de apoyo al aprendizaje profundo.
- **Pytorch [27]:** PyTorch es compatible con las principales plataformas en la nube y permite fácil escalado a través de imágenes preconstruidas. Permite exportar modelos de formato estándar ONNX (Open Neural Network Exchange) para acceder directamente a plataformas también posee visualizadores compatibles con ONNX. Tiene soporte nativo para la ejecución asíncrona de operaciones colectivas y la comunicación entre pares para acceder desde Python y C++.
- **RapidMiner [28]:** Es una herramienta inteligente con una interfaz amigable para los no programadores que permite el manejo de datos para la preparación de la fase de estos en la fase de entrenamiento, y es fácil de utilizar para luego generar el modelo gracias a AutoModel, además, posee independencia de la plataforma, ya que funciona en sistemas operativos multiplataforma.
- **Google Cloud AutoML[29]:** Proporciona los modelos previamente entrenados para los usuarios a fin de crear diversos servicios.
- **Jupyter Notebook [30]:** Es un procesamiento muy rápido, así como una plataforma eficiente. Su nombre se conforma por estos lenguajes de programación: Julia, R, Python.
- **Apache Mahout [31]:** Mahout es parte de Apache, que es una plataforma de código abierto basada en Hadoop.
- **Estudio Azure Machine Learning [32]:** Este es el producto de Microsoft que proporciona servicios de aprendizaje automático, Azure Machine Learning Studio es una forma muy fácil de formar conexiones de módulos y conjuntos de datos.
- **Orange3[33]:** Exploración interactiva de datos para un análisis. La interfaz gráfica de usuario le permite concentrarse en el análisis exploratorio de datos en lugar de la codificación, mientras que los valores predeterminados inteligentes hacen que la creación rápida de prototipos de un flujo.

3 Análisis propuesto de evaluación

Se utilizó el conjunto de datos llamado “*Heart Disease Dataset*” [34], este es un conjunto de datos de enfermedades del corazón, el cual presenta un registro simple de pacientes cardíacos monitoreados con problemas de clasificación binaria. Este dataset se encuentra disponibilizado en el sitio web de Kaggle [35], que es una subsidiaria de Google LLC. Esta es una comunidad en línea de científicos de datos y profesionales del aprendizaje automático, la cual cuenta con varios conjuntos de datos públicos especialmente utilizados para machine learning, cursos, foros, prácticas entre colegas y demás.

En base al análisis de herramientas relevadas en la sección anterior, por su simplicidad, amigabilidad en la utilización de funciones del software y por su soporte en foros de usuarios, se utilizó la herramienta Orange3 [33] de aprendizaje automático de código abierto y visualización de datos, en el cual se desarrollan los flujos de trabajo de análisis. En la Figura 1, se muestra el esquema propuesto a modo de análisis.

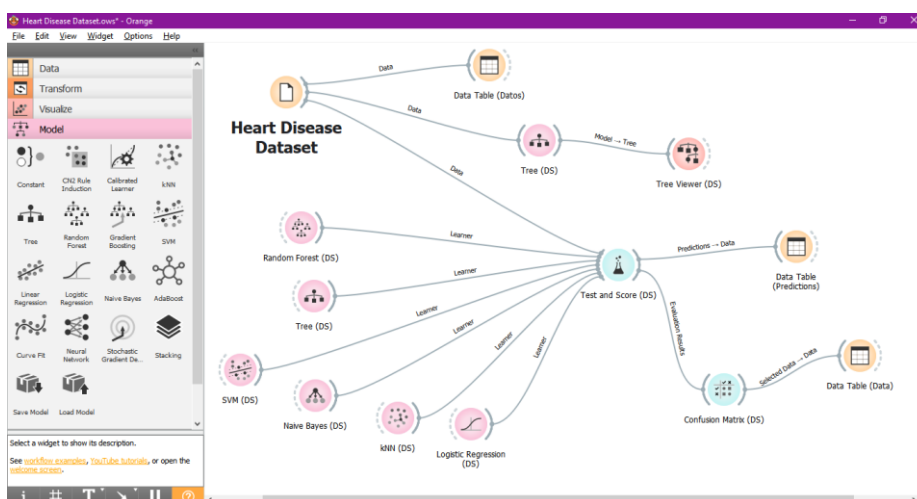


Fig. 1. Se muestra el diagrama con nodos de las funciones utilizadas para el análisis de tests de tipos de técnicas de machine learning.

Como se observa en la Figura 1, inicialmente se conectó una fuente de datos (dataset seleccionado, elemento “*File*”), luego se incorporó el elemento de evaluación “*Test and Score*”, el cual fue conectado con los distintos modelos, siendo éstos: a) Random Forest; b) Tree; c) SVM; d) Naive Bayes; e) kNN; f) Logistic Regression. Cada uno de los modelos mencionados anteriormente fue analizado a través de una evaluación de resultados (“*Confusion Matrix*”) en la siguiente sección, se muestran los resultados arrojados.

Este es un elemento de evaluación en Orange, el cual visualiza los datos que no fueron clasificados correctamente y la que sí, por ello se denomina, matriz de confusión (se muestra el resultado final del algoritmo de clasificación para aprendizaje

supervisado). Los resultados arrojados en dichos modelos se muestran en una tabla comparativa con los distintos enfoques utilizados.

El elemento de evaluación “*Matriz de confusión*”, muestra las proporciones entre la clase predicha y la real. La selección de los elementos en la matriz alimenta las instancias (muestra las instancias clasificadas correcta e incorrectamente numéricamente) correspondientes en la señal de salida. De esta manera, se puede observar qué instancias específicas fueron mal clasificadas y cómo [36].

Por otra parte, se agregó un modelo, elemento árbol de decisión con el fin de visualizar los grupos que se analizan desde un nodo inicial y en base al valor de una variable, representar como se va bifurcando en las “hojas” del árbol (elemento “Tree Viewer”), esto se muestra en la Figura 2.

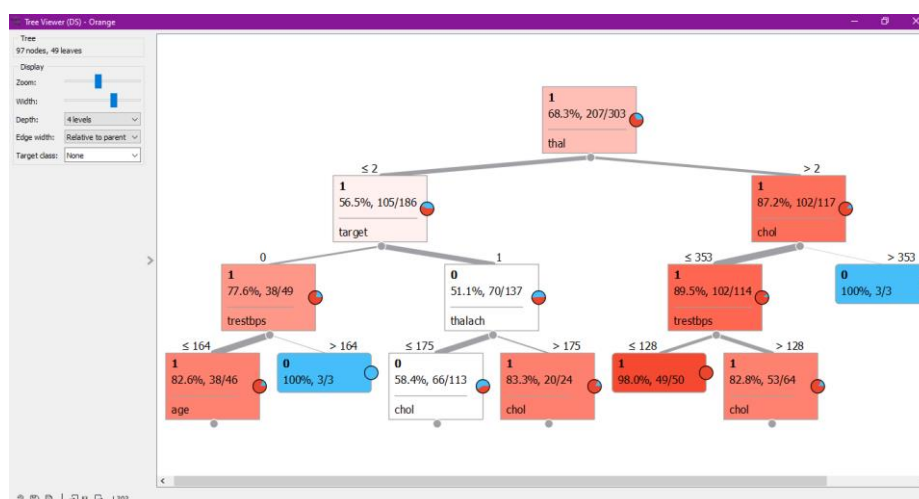


Fig. 2. Se muestra el diagrama de árbol de decisión de 4 niveles, con sus caminos.

Cabe destacar que en la Figura 2, los niveles de visualización de los datos se pueden configurar, es decir, en este ejemplo se visualizan 4 niveles.

En la siguiente sección, se muestran los resultados arrojados luego del análisis de los modelos relevados.

4 Resultados

La aplicación Orange tiene un widget que obtiene los resultados de la evaluación de “*Test & Score*” (prueba y puntuación, es decir prueba algoritmos de aprendizaje sobre datos), esto se muestra en la Figura 3.

En la Figura 3, se observa que se toma un muestreo aleatorio y le indicamos que lo entrenamos con el 70% de los datos tomados del archivo que es lo recomendado para estos casos de análisis. Para este análisis se indica que el modelo que presenta mayor precisión es el SVM con 0.723, seguido por el Naive Bayes con 0.700, esto quiere decir que son los modelos con la proporción de verdaderos positivos entre instancias

8

clasificadas como positivas más altos para este dataset, y se observa que el que tiene una menor precisión es kNN.

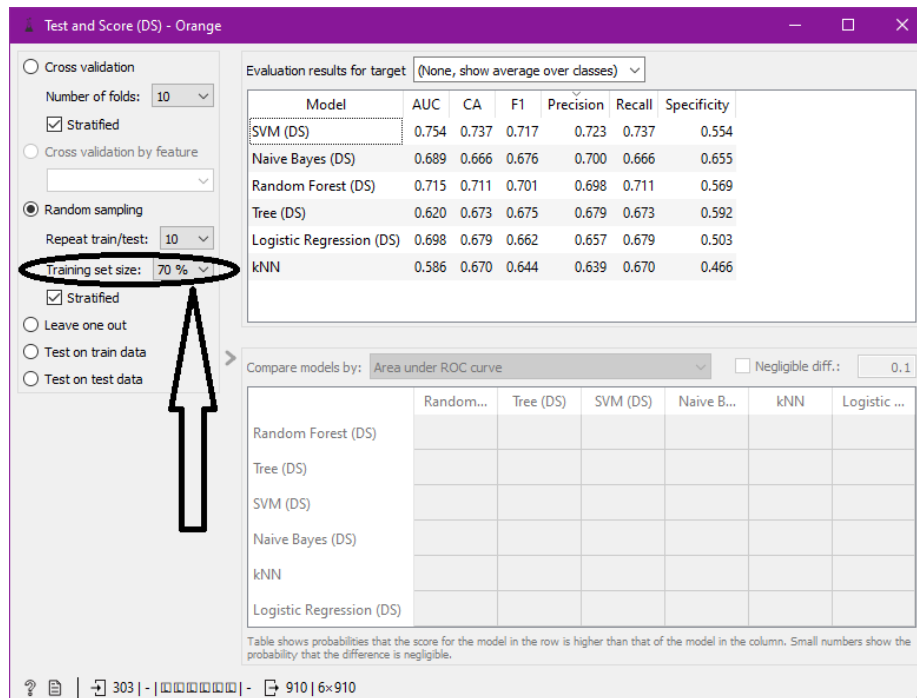


Fig. 3. Se muestra la tabla de “Test and Score” con los resultados arrojados por los distintos modelos analizados.

A modo de resumen, el cuadro identifica algunos aspectos siendo estos:

- **AUC**=El área bajo ROC es el área bajo la curva de funcionamiento del receptor.
- **CA**=La precisión de la clasificación es la proporción de ejemplos correctamente clasificados.
- **F-1**=es una media armónica ponderada de precisión y recuperación.
- **Precision**=La precisión es la proporción de verdaderos positivos entre instancias clasificadas como positivas, por ejemplo, la proporción de Tipo de sexo femenino correctamente identificado como Tipo de sexo femenino.
- **ReCall**=El recuerdo es la proporción de verdaderos positivos entre todas las instancias positivas en los datos, por ejemplo, el número de enfermos entre todos los diagnosticados como enfermos.
- **Specificity**=La especificidad es la proporción de verdaderos negativos entre todos los casos negativos, por ejemplo, el número de no enfermos entre todos los diagnosticados como no enfermos.

Si analizamos los resultados arrojados en el widget “*Confusion Matrix*”, se visualiza cada uno de los modelos tratados, junto a una tabla con la proporción de predicción, en este caso, en la Figura 4 (ventana izquierda), se muestran los porcentajes correctos/acertados, por ejemplo, para el caso del tipo 0, se predijo el valor cero con un 64.4% de aciertos, lo mismo que para el caso 1 con la predicción correcta 1 en un 76%, si se observa el caso en color rojo, para el valor 1, se predijo 0 (erróneamente con una proporción de 35.6%), lo mismo sucede para el valor 0, se predijo 1 (erróneamente con una proporción de 24%).

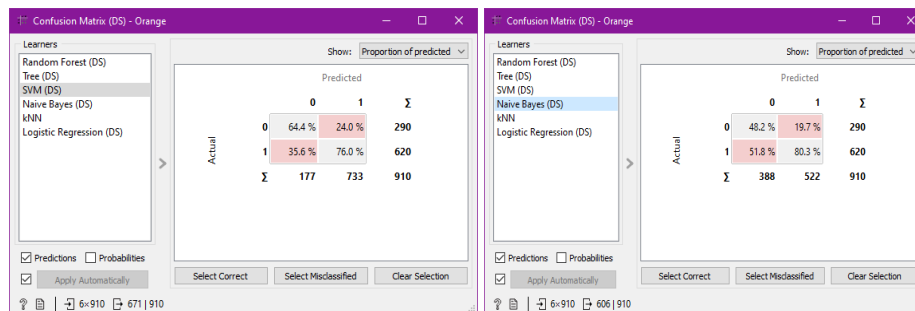


Fig. 4. Se muestra la ventana del software Orange con la sección de “*Confusion Matrix*” para el modelo SVM (ventana izquierda) y modelo (ventana derecha).

En la Figura 4 (ventana derecha), se muestran los porcentajes correctos/acertados, por ejemplo, para el caso del tipo 0, se predijo el valor cero con un 48.2% de aciertos, lo mismo que para el caso 1 con la predicción correcta 1 en un 80.3%, si se observa el caso en color rojo, para el valor 1, se predijo 0 (erróneamente con una proporción de 51.8%), lo mismo sucede para el valor 0, se predijo 1 (erróneamente con una proporción de 19.7%).

5 Trabajos futuros y Conclusiones

Como propuesta de trabajo futuro, los autores realizarán nuevos tests con este modelo propuesto de análisis para un mayor conjunto de datos públicos abiertos y de esta manera analizar el algoritmo de mejor precisión en base a una muestra determinada.

Esta investigación aportó un relevamiento de las herramientas más utilizadas en la actualidad y, además, brindó una breve descripción de algunos de los algoritmos de aprendizaje del tipo supervisado. Cabe aclarar que si bien cada algoritmo de aprendizaje tiene un enfoque distinto puede ser fusionado o utilizado en conjunto con otro algoritmo para poder obtener mejores resultados a nivel de predicción. Este trabajo muestra la forma de evaluación de los algoritmos a través de la utilización de la herramienta Orange, la cual facilitó la evaluación por su entorno amigable y visual, pero no forma parte de este estudio los casos de predicción, ya que esto conllevaría a una limpieza previa del dataset a ser utilizado o bien estudiado.

Por otra parte, si bien el concepto de machine learning se encuentra en pleno auge es sumamente importante concientizar y educar a los ciudadanos en estas tecnologías, pero hay un reto aún mayor en lo que nos depara este largo camino y es lograr una convergencia de los datos públicos abiertos como así también el análisis de los datos privados que deben alimentar a este tipo de tecnologías con el fin de generar un valor agregado a los ciudadanos tanto a nivel nacional como internacional.

Bibliografía

1. César, C., & Lorenzo, S. (2010). Open government: gobierno abierto. Jaén, España: Algón Editores MMX, 2010.
2. García, S. Á., Casado, M. C. G., & Barrio, M. G. (2021). Comunicación digital basada en datos abiertos y la Agenda 2030 en Latinoamérica: nuevas oportunidades. *Comunicación y sociedad*, 1-20.
3. Martínez, R. (2022). Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales (Doctoral dissertation, Universidad Nacional de La Plata).
4. Janiesch, C., Zszech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695.
5. Datos.gob.ar, “¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos”, <https://datos.gob.es/es/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos>, último acceso 06/2022.
6. Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018, March). Machine learning in healthcare: A review. In 2018 Second international conference on electronics, communication and aerospace technology (ICECA) (pp. 910-914). IEEE.
7. Galvão, T. L., Ferreira, I., Kuznetsova, A., Novell-Leruth, G., Song, C., Feiler, C., ... & Tedim, J. (2022). CORDATA: an open data management web application to select corrosion inhibitors. *npj Materials Degradation*, 6(1), 1-4.
8. Marqués Gozalbo, M. Á. (2022). Modelos predictivos de producción agroindustrial con Machine Learning a partir de fuentes de información pública.
9. Fernández Caraballo, E., & Gómez Franco, Y. (2018). Metodología para el análisis de la violencia en el departamento de Bolívar mediante técnicas de machine learning.
10. Chamorro, K., Laza, N., Noriega, H., Rojano, R., Vega, J., & Heredia, D. (2021). Aplicación de Machine Learning para análisis de los fenómenos de violencia intrafamiliar en el departamento del Atlántico. *Investigación y desarrollo en TIC*, 12(1).
11. Herrera Carrasco, J. (2020). Evaluación de Modelos de Transporte mediante datos abiertos y técnicas de Aprendizaje Automático.
12. Saravanan, R., & Sujatha, P. (2018, June). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 945-949). IEEE.
13. Zhang, S., Li, X., & Zong, M. (2017). Learning k for kNN Classification. *ACM Transactions on Intelligent Systems and Technology*, pp 1-19.
14. Zhang, C., Zhong, P., Liu, M., Song, Q., Liang, Z., & Wang, X. (2022). Hybrid Metric K-Nearest Neighbor Algorithm and Applications. *Mathematical Problems in Engineering*, 2022.
15. William, S. (2018). Decision Tree and Random Forest - Machine Learning and Algorithms: The Future Is Here! United Kingdom: WorldofBooks.

16. Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. JASTT, Vol. 02, No. 01, pp. 20 – 28.
17. Tu, P., & Chung, J. (1992). A new decision-tree classification algorithm for machine learning. Proceedings Fourth International Conference on Tools with Artificial Intelligence, pp. 370,371,372,373,374,375,376,377. doi:10.1109/TAI.1992.246431
18. IBM, “Funcionamiento de SVM”, <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=models-how-svm-works>, último acceso 06/2022.
19. itoo F, M. s. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. International Journal of Information Technology, 1503-1511. doi:10.1007/s41870-020-00430-y
20. Jason, B. (2019). Master Machine Learning Algorithms: Discover How They Work and Implement them from scratch. Master Machine Learning algorithm.
21. Das, S. K., Das, S. P., Dey, N., & Hassanien, A.-E. (2020). Machine Learning Algorithms for Industrial Applications. Springer Nature.
22. Knime, “End to End Data Science”, <https://www.knime.com/>, último acceso 06/2022.
23. Accord-framework.net, “Machine learning made in a minute”, <http://accord-framework.net/>, último acceso 06/2022.
24. Scikit-learn, “Machine Learning in Python”, <https://scikit-learn.org/stable/>, último acceso 06/2022.
25. TensorFlow, “Plataforma de extremo a extremo de código abierto para el aprendizaje automático”, <https://www.tensorflow.org>, último acceso 06/2022.
26. Weka, “Welcome to a New Era”, <https://www.weka.io/>, último acceso 06/2022.
27. Pytorch, “End-to-end Machine learning Framework”, <https://pytorch.org/features/>, último acceso 06/2022.
28. Rapidminer, “People x Expertise x Data”, <https://rapidminer.com/>, último acceso 06/2022.
29. Cloud.google, “AutoML”, <https://cloud.google.com/automl?hl=es>, último acceso 06/2022.
30. Jupyter, “Jupyter”, <https://jupyter.org/>, último acceso 06/2022.
31. Mahout, “Mahout”, <https://mahout.apache.org/>, último acceso 06/2022.
32. Azure Machine Learning, “Azure Machine Learning”, <https://azure.microsoft.com/es-es/services/machine-learning/#features>, último acceso 06/2022.
33. Orange, “Data Mining Fruitful and Fun”, <https://orangedatamining.com/>, último acceso 06/2022.
34. Kaggle, “Heart Disease Dataset”, <https://www.kaggle.com/datasets/yasserh/heart-disease-dataset?resource=download>, último acceso 06/2022.
35. Kaggle, “Start with more than a blinking cursor”, <https://www.kaggle.com/>, último acceso 06/2022.
36. Orange, “Confusion Matrix”, <https://orangedatamining.com/widget-catalog/evaluate/confusionmatrix/>, último acceso 06/2022.