

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325860155>

Machine learning: Importance, advances, techniques and applications

Chapter · April 2018

CITATIONS

0

READS

85

3 authors, including:



Luis Suárez Zambrano

Universidad Técnica del Norte

21 PUBLICATIONS 25 CITATIONS

[SEE PROFILE](#)



Omar Oña

Universidad de las Fuerzas Armadas-ESPE

15 PUBLICATIONS 11 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



METODOLOGÍA DE VISUALIZACIÓN INTERACTIVA Y ANÁLISIS DE INFORMACIÓN EN BIG DATA. [View project](#)



Factors that influence the teaching / learning of mathematics in engineering [View project](#)



ANA CRISTINA UMAQUINGA-CRIOLLO

Estudiante de doctorado en la Universidad de Salamanca - España, estudiante invitado en la Universidad de Salamanca – España. Magister en Gerencia Informática, Ingeniera en Sistemas Computacionales.

Experiencia académica en docencia e investigación y experiencia empresarial en soporte de tecnología de la información para empresas bancarias.

Experiencia docente: programación, multimedia, base de datos, informática general, producción de audio y video, graficación y animación.

Experiencia empresarial: 11 años de experiencia en soporte de TI para IBM Ecuador y "Banco del Pacífico - Ibarra"

Áreas de investigación: Tecnología de la información, sector bancario, Big Data, Big Data Visualization, Moodle Virtual Classrooms. Ingeniería y Tecnología, Machine learning



LUIS EDILBERTO SUÁREZ-ZAMBRANO

Magíster en Gerencia Informática por la Pontificia Universidad Católica del Ecuador Sede Ibarra, Ingeniero en Sistemas Computacionales por la Universidad Técnica del Norte. Ha desempeñado como Docente Tiempo Completo en la Universidad Regional Autónoma de Los Andes UNIANDES-IBARRA 2011-2015, en la carrera de Sistemas e Informática, cargo como Coordinador de Vinculación con la Sociedad 2013-2015.

Actualmente desempeña como Docente Titular Auxiliar en la Facultad de Ingeniería en Ciencias Aplicadas FICA, de la Universidad Técnica del Norte.

Docente Investigador en áreas de Machine Learning, Big Data, Aprendizaje Automático, Data Mining, Big Data Visualization, Dimensionality Reduction.



OMAR RICARDO OÑA-ROCHA

Ingeniero en Electrónica y Telecomunicaciones por la Universidad de las Fuerzas Armadas (ESPE, Quito-Ecuador.

Actualmente se desempeña como profesor en la Facultad de Ingeniería en Ciencias Aplicadas (FICA), carrera de Ingeniería en Telecomunicaciones (CITEL) de la Universidad Técnica del Norte, Ibarra-Ecuador.

Docente investigador en áreas de Matemática, Currículo, Big Data, Machine Learning, Electrónica y Telecomunicaciones.

Pasto, 19-20 de abril del 2018

Ana Umaquina, Luis Suárez, Omar Oña (Universidad Técnica del Norte / Ibarra-Ecuador)

El aprendizaje automático: Importancia, avances, técnicas y aplicaciones

Machine learning: Importance, advances, techniques and applications

(Entregado 23/03/2018 – Revisado día/mes/año)

Universidad Técnica del Norte

Ana Cristina Umaquina-Criollo, Luis Edilberto Suárez-Zambrano, Omar

Ricardo Oña-Rocha

Facultad de Ingeniería en Ciencias Aplicadas - Carrera de Ingeniería en

Telecomunicaciones

Ibarra – Ecuador

acumaquina@utn.edu.ec, lesuarez@utn.edu.ec, oronia@utn.edu.ec

593-062997800 Ext: 7506

Resumen

Este trabajo presenta un estudio descriptivo de las técnicas de aprendizaje automático llamado machine learning, enfocándose en las técnicas de aprendizaje supervisado y no supervisado, con la presentación de las técnicas exitosas para determinados casos de estudio, experimentación y puesta en producción.

Esta investigación pretende ser de aporte a la comunidad científica y empresarial, principalmente investigadores que inician y quienes experimentan sus estudios en análisis comparativos en esta área de vital importancia.

Este trabajo es de tipo investigativo, documental y exploratorio. Se realiza un recorrido por los principales tipos de aprendizaje automático, avances y proyectos enfocados a Inteligencia Artificial.

Pasto, 19-20 de abril del 2018

Ana Umaquina, Luis Suárez, Omar Oña (Universidad Técnica del Norte / Ibarra-Ecuador)

Página 2 de 17

Palabras clave: *Aprendizaje Automático, Aprendizaje Supervisado, Aprendizaje no-Supervisado, Inteligencia Artificial, Aplicaciones de Aprendizaje Automático.*

Abstract

This paper presents a descriptive study of machine learning techniques called machine learning, focusing on supervised and unsupervised learning techniques, with the presentation of successful techniques for specific cases of study, experimentation and production.

This research aims to be of contribution to the scientific and business community, mainly researchers who initiate and who experience their studies in comparative analysis in this area of vital importance.

This work is of investigative, documentary and exploratory type. A tour through the main types of machine learning, advances and projects focused on Artificial Intelligence.

Keywords: *Machine Learning, Supervised Learning, Unsupervised Learning, Artificial Intelligence, Machine Learning Applications.*

1. Introducción

Ante el crecimiento vertiginoso e innegable de los datos tanto en tamaño y tipo: estructurados, semi estructurados, no estructurados, principalmente, produce sobrecarga de información, de ahí que, entre los principales desafíos de las áreas, tecnológicas , investigación y empresarial se encuentra el descubrimiento , análisis, y tratamiento de información (Umaquinga, Peluffo, Alvarado P., & Cabrera A., 2016), y que permite ser

visualizada de una manera entendible para los usuarios o clientes finales (IBM, 2012), (Umaquina et al., 2016).

“La inteligencia artificial, aprendizaje automático y minería de datos son herramientas que facilitan el análisis de datos” (Alvarado-Pérez, Peluffo-Ordóñez, & Therón, 2015). El Machine Learning (Aprendizaje Automático) busca desarrollar técnicas para que las “computadoras aprendan”, (Alvarado-Pérez et al., 2015), con técnicas entre ellas, las que con el apoyo de expertos o clasificadores supervisados previamente conocidos para clasificar a un nuevo elemento dentro del universo o conjunto de entrenamiento.

El documento se organiza de la siguiente manera. En la sección 2 se realiza un breve recorrido iniciando por la definición de Aprendizaje Automático para introducir al lector en el tema de estudio de las subsecciones 2.1, 2.2 en donde se analiza las propuestas de clasificación en aprendizaje supervisado y no supervisado, y experimentaciones exitosas de la aplicación algunas propuestas de este tipo de aprendizaje, en la subsección 2.3 se presentan los principales aplicaciones. La sección 3 presenta las conclusiones, finalmente en la sección 4 presenta las recomendaciones obtenidas en este estudio.

2. APRENDIZAJE AUTOMÁTICO

El Aprendizaje automático o Machine learning es una rama dentro de la Inteligencia Artificial, que comprende el aprendizaje a partir de la experiencia (Sammut & Webb, 2011), que se basa en la construcción de programas computacionales que automáticamente mejoren su rendimiento en una tarea determinada en base a la experiencia.

“Un programa de computadora se dice que aprende de experiencia **E** con respecto a una clase de tareas **T** y medida de desempeño **D**, si su desempeño en las tareas en T, medidas con D, mejoran con experiencia” (Mitchell, 1997)

Dentro del subárea del reconocimiento de patrones (Pattern recognition) existen 2 propuestas para tratar el tema de la clasificación: Método de clasificación supervisada y no supervisada. (Araujo, 2006)

2.1.Técnicas de Clasificación supervisadas

Para el método supervisado, el conjunto de objetos descritos por un vector de características y la clase de pertenencia en el conjunto de entrenamiento, ésta técnica obtiene un modelo o regla general para clasificarlos (Araujo, 2006)

Dentro del aprendizaje inductivo es la “capacidad de obtener nuevos conceptos, más generales, a partir de ejemplos”. Existen 2 tipos de datos para el conjunto de datos: entrenamiento y prueba que se utilizan en el conjunto de entrenamiento.

2.1.1. Clasificación

Entre las principales clasificaciones dentro de data mining se encuentran (Alvarado-Pérez et al., 2015):

Clasificación:

- Tablas de decisión. Decision tables
- Árboles de decisión. Decision Trees
- Reglas Bayesianas. Bayesian rules
- Redes neuronales. Neural Networks
- Algoritmos genéticos. Genetic Algorithms
- Lógica difusa. Fuzzy logic

Predicción:

- Regression
- Trees Prediction

- Core estimator

Para el área de aprendizaje supervisado, las técnicas utilizadas son:

- K-NN
- Regresión: lineal, polinomial
- Clasificación
- Naive Bayes
- pre-procesamiento
- Árbol de decisión

2.1.2. Los algoritmos de predicción por vecindad:

Se basan en argumentos heurísticos, es decir de la experiencia. La clasificación de un nuevo punto o caso X dentro de la categoría, espacio, o dominio de una colección de N casos, conjunto de entrenamiento previamente clasificados se calcula en función de (Araujo, 2006):

- Las clases conocidas y clasificadas de antemano $(x_1, \theta_1), (x_2, \theta_2), \dots, (\theta_N, \theta_N)$.
- Los puntos más cercanos a él conocidos previamente.
- Basan su efectividad en la utilización de una métrica adecuada

2.1.3. K-nn (K-vecinos más cercanos), del inglés K-Nearest Neighbors

Es uno de los modelos de clasificación propuestos en el área de aprendizaje automático, se basa en las distancias. El parámetro K determina el número de los vecinos más cercanos para predecir sobre la clase

En función de la distancia:

- Distancia euclídea: Distancia de la línea recta

$$d(x, z) = \sqrt{\sum_{i=1}^d (x_i - z_i)^2} \quad (1)$$

- Distancia euclídea cuadrada: Elimina la raíz cuadrada de la ecuación
- Distancia Manhattan: Distancia entre 2 puntos tomando como base una cuadrícula

$$d(x, z) = \sum_{i=1}^d |x_i - z_i| \quad (2)$$

- Distancia Chebshev: Distancia con la mayor diferencia entre las características de ambos vectores

$$d(x, z) = \max_i |x_i - z_i| \quad (3)$$

2.1.4. Variantes adicionales de K-nn:

- **Regla de K-nn con rechazo:** Descarta muestras que no tengan garantía de que la clase asignada sea correcta
- **Regla K-nn por distancia media:** Da mayor relevancia a la distancia a la que se encuentran los K vecinos más próximos que al voto individual de cada vecino
- **Distancia mínima**

El éxito de esta técnica depende de la correcta selección de la función de distancia a aplicar precisa de la experiencia y del conocimiento del experto en relación al tema que se clasifica.

Estudios de Clasificación de células cervicales mediante el algoritmo KNN usando rasgos del núcleo distinguen aproximaciones de células normales y anómalas utilizando las distancias: Mahalanobis, euclidiana y Manhattan, siendo la distancia Manhattan la más

adecuada a utilizar para la clasificación de células cervicales (Rodríguez Vázquez, Martínez Borges, & Lorenzo Ginori, 2015).

Existen estudios comparativos en casos prácticos de estudio de bases de datos de enfermedades cardiacas (Cleveland), cardiocografía e hipotiroidismo que toman clasificadores recomendados en la literatura científica, como: El clasificador de K vecinos más cercanos K-NN, ANN, máquinas de vectores de soporte (SVM), el clasificador Parzen (PC), árboles de decisión (Random Forest) y el clasificador Adaboost (Basante, Ortega, Peluffo, & Blanco, 2017).

2.2. Técnicas de Clasificación no supervisadas

El aprendizaje no supervisado, es un método del Aprendizaje Automático, que abarca un conjunto de técnicas, algoritmos y secuencias, que permite agrupar datos de forma rápida; conocidos también como métodos simétricos o indirectos, los “*datos observados*” se clasifican en grupos o clases que no pueden ser determinados previamente, éstas clases o grupos de variables pueden o no estar vinculadas entre sí y todas las observaciones son evaluadas a un mismo nivel (Emtiyaz & Keyvanpour, 2011). En este tipo de aprendizaje no se requiere la intervención del ser humano para la categorización de las observaciones, por cuanto los datos no se encuentran “*etiquetados*”, es tarea del algoritmo encontrar patrones dentro de los datos considerando su distribución y composición y clasificarlos para posterior entrenamiento del algoritmo. Una de las áreas que abarca el aprendizaje no supervisado se denomina técnicas de Clustering, que son tareas cuyo propósito es agrupar conjuntos de datos no etiquetados en subconjuntos denominados Clusters. Cada Cluster está formado por una colección de objetos de similares características, pero que son distintos respecto a los datos de otros Clusters. (Moya, 2016). La tarea fundamental del Clustering es ayudar a entender y resumir la data.

El aprendizaje no supervisado supone un mecanismo de ayuda para la toma de decisiones en diferentes situaciones en las que el ser humano difícilmente las podría tomar de manera correcta debido a las características muchas veces ocultas de los datos recolectados de un

determinado fenómeno o circunstancias observados. Depende en gran porcentaje del método o algoritmo de agrupamiento utilizado para obtener mejores resultados en cuanto a la clasificación correcta de los datos y a la futura toma de decisiones. Así mismo, este tipo de aprendizaje es útil para ayudar a comprender los datos y se resumen o derivan en una distribución de probabilidad sobre un conjunto de datos de entrada.

En la actualidad se tienen una variedad de técnicas, métodos y algoritmos para para aprendizaje no supervisado, entre ellos se pueden mencionar a las **Técnicas de aprendizaje no supervisado**: Clustering, Visualización, Reducción de la dimensionalidad (DR), Extracción de características, entre otras. (Ng, Jordan, & Weiss, 2011) **Algoritmos para Clustering**: K-Means Clustering, Mezclas Gaussianas, Clustering Jerárquico Aglomerativo, Density-based spatial clustering of applications with noise (DBSCAN): Basado en densidades, Redes neuronales, entre otros. (Castro, Castro, Peluffo, & Castellanos, 2011). Con el análisis de datos no supervisado, encontramos varias características válidas tales como: Escalabilidad, clusters de forma arbitraria, manejo de diferentes tipos de atributos, la capacidad de añadir restricciones, manejo de ruido o datos erróneos, funcionan eficientemente con altas dimensionalidades de los datos, se necesitan mínimos requerimientos cuando se requiere especificar parámetros como por ejemplo el número de clusters, independencia del orden de los datos, clusters interpretables y manejables. (Cambronero & Moreno, 2006).

- **K-Means Clustering**: es tal vez uno de los algoritmos más representativos en el campo del Aprendizaje Automático con enfoque en el clústering. Dado un N conjunto de instancias de D dimensiones, el objetivo de este algoritmo es clasificar o particionar los datos en k clusters (para un k dado) y seleccionar al azar k centroides desde μ_1, \dots, μ_k y repetir el proceso hasta que los centroides no varíen de tal manera que se minimicen los datos u observaciones que no correspondan a un cluster.

$$\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \quad (4)$$

El algoritmo tiene como ventajas el hecho de ser simple en su trabajo, eficiente en cuanto a la clasificación de los datos y de propósito general en cuanto a aplicación

en cualquier área de investigación. Pueden presentarse desventajas tales como la necesidad de especificar k o el hecho de que no siempre puede solucionar los problemas con múltiples inicializaciones.

- **Mezclas Gaussianas:** se constituye en un modelo probabilístico que pretende buscar sub clusters dentro de un mismo clúster y que representa generalmente a una distribución de probabilidad de mezcla de alguna observación o dato en la totalidad de los datos observados.

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (5)$$

La idea principal en este algoritmo es que cada clúster se ajuste a una distribución normal en donde los datos puedan ajustarse a un modelo de mezclas de Gaussianas (GMM).

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (6)$$

En donde π_k son los coeficientes de mezcla, que cumplen:

$$0 \leq \pi_k \leq 1 \quad \text{y} \quad \sum_{k=1}^K \pi_k = 1 \quad (7)$$

El campo de aplicación de este algoritmo es muy amplio, desde modelar diferentes grupos etarios, étnicos, la estatura de las personas, pasando por reconocimiento de patrones en el habla del ser humano, fonemas, hasta texturas de patologías en la salud humana.

- **Clustering Jerárquico Aglomerativo:** este algoritmo consiste en definir un cluster por cada instancia de datos y fusionar los clusters más cercanos entre si hasta que quede un único clúster, aunque supone un grado de complejidad mayor por el número de instancias que hay que calcular o encontrar. La distancia entre los clusters es el fundamento del algoritmo, y puede variar debido a la forma de los clusters que el algoritmo puede manejar, para lo cual se necesitar definir *dendogramas* muy útiles para crear taxonomías por ejemplo. Una de las desventajas del algoritmo es que supone un alto coste computacional.

- **DBSCAN (Density-based spatial clustering of applications with noise):** como su nombre lo indica, este algoritmo es ideal para escenarios en donde los datos provienen de grandes bases de datos como los espaciales por ejemplo, es un algoritmo de agrupamiento de datos cuyo objetivo consiste en modelar clusters de alta densidad de datos, en donde si un dato pertenece o no a un clúster éste debe estar cerca de un gran número de datos de dicho clúster. Es un algoritmo de mucha utilidad cuando se tiene formas irregulares, hay ruido, o los datos se encuentran entrelazados.
- **Redes Neuronales:** si bien las redes neuronales pueden usarse como método o técnica de aprendizaje supervisado, su mayor aplicación lo es en escenarios de aprendizaje no supervisado. Existen varios tipos de redes neuronales en la actualidad entre ellas tenemos:
 - El Perceptron Simple.
 - La Red de Hopfield.
 - El Perceptron Multicapa.
 - Red neuronal Competitiva Simple.
 - Redes Neuronales Online ART1.
 - Redes Neuronales competitivas ART2.
 - redes neuronales autoorganizadas: Mapas de Kohonen.

2.3. Aplicaciones de Aprendizaje Automático

El recorrido histórico del aprendizaje automático que surgió en los años 80-90 como un planteamiento integrador de los enfoques de la Inteligencia artificial (IA) la que se ocupa básicamente de aspectos más cognitivos estrictamente relacionados con la lógica, el conocimiento y su procesamiento y del Reconocimiento de formas (RF) o reconocimiento de patrones que se ocupa de aspectos más perceptivos y relacionados con la visión, el habla, etc.

En la actualidad el aprendizaje automático prácticamente cuenta con tantas aplicaciones como se pueda imaginar, pudiéndose adaptar a tantas situaciones como datos con los que

Pasto, 19-20 de abril del 2018

Ana Umaquinga, Luis Suárez, Omar Oña (Universidad Técnica del Norte / Ibarra-
Ecuador)

se cuente. Por lo que se puede citar muchas áreas en las que se desarrolla una gran variedad de aplicaciones, como son:

Reconocimiento de imágenes: Análisis de imágenes de alta calidad, reconocimiento de caracteres, de manuscritos, firmas, piezas industriales, texturas, tipos de vehículos, identificación de placas, detección de defectos en control de calidad, etc.

Robótica y Visión artificial: Detección y reconocimiento de rostros y expresiones faciales, visión para navegación de robots, vehículos autónomos, exploración espacial, etc.(p.e. en China, gracias al reconocimiento facial la policía utiliza gafas que permiten identificar a cualquier persona en medio de una multitud ya que pueden procesar hasta 10000 caras en una décima de segundo disminuyendo en alto grado los actos delictivos, el gobierno espera identificar a cada uno de los 1300 millones de ciudadanos en apenas 3 segundos y con un porcentaje de acierto de al menos el 88%) (Fontdeglòria, 2018).

Reconocimiento del habla y procesamiento de lenguaje: Comprensión y traducción automática, reconocimiento de palabra aisladas, habla continua, etc.

Predicción y pronósticos: De clima, tráfico o para evitar fallos tecnológicos en equipos, de enfermedades, etc.

Aplicaciones de Teledetección: Usando imágenes aéreas o de satélite se centra en la exploración de recurso naturales, predicción de cosechas y explotaciones forestales, localización de posibles yacimientos minerales, etc.

Aplicaciones en Biomedicina y Genómica: Detección de tumores y tejidos cancerosos, detección de situaciones críticas en las unidades de cuidados intensivos, análisis de electro-cardio-encefalogramas (señales ECG), diagnósticos a partir de síntomas, reconocimiento de cromosomas para detección de malformaciones congénitas, clasificación de secuencias de ADN, análisis de secuencias genómicas.

Análisis de señales sísmicas: Dentro de señales naturales, la predicción de terremotos; señales artificiales la localización de yacimientos de minerales y petróleo.

Aplicaciones Biométricas: Reconocimiento de huellas dactilares, de rostros, de iris, análisis de voz, etc.

Aplicaciones Agrícolas: Mediante la visión artificial para recolección automática, localización de malas hierbas y su eliminación de manera selectiva, detección de puntos de injerto y lograr la automatización, selección y envasado de frutos y otros.

Análisis de comportamiento de consumo y productividad: La identificación de clientes potenciales, prever qué empleados pueden ser más rentables, adaptar servicios a las necesidades del usuario.

Aplicaciones económicas: Predicciones de tendencias de patrones de fraude, segmentación de clientes, previsión de ventas, minería y análisis de datos y analítica de riesgos.

Detección de correos molestos (SPAM): Clasificar aquellos correos que no tienen absolutamente nada que hacer en tu bandeja de entrada mediante el uso de etiquetas.

Motores de búsqueda: Hoy tienen la habilidad de aprender de los hábitos y costumbres de los usuarios, de manera que permite a los productos acercar información relevante en el momento preciso.(p.e. “Según Lori Lewis y Chadd Callahan, infógrafos de Cumulus Media, solo en un minuto de internet se envían 156 millones de Emails, 900.000 personas se conectan a Facebook, 3.5 millones de usuarios realizan búsquedas en Google, se envían 452.000 tuits, se suben 46.200 fotos a Instagram, se visualizan 70.017 horas de contenido en Netflix y se reproducen 4,1 millones de videos en Youtube”)

<https://www.elspectador.com/tecnologia/asi-va-la-inteligencia-artificial-y-el-aprendizaje-automatico-de-google-articulo-719816>

Protección Civil: Detección de casos de alerta en sistemas hidrológicos, predicción del clima, control de incendios forestales, predicción de movimientos telúricos, etc.

Aplicaciones para discapacitados: Ayudas para la visión, para el aprendizaje del habla, control del entorno mediante reconocimientos del habla, etc.

El machine learning o el aprendizaje automático transformarán el mundo tal como lo hizo el motor a finales del siglo XIX”, aseguró Edgar Helou, director de Google Cloud Colombia, en medio de una conferencia sobre el presente y el futuro de estas tecnologías y su efecto en la vida cotidiana de las personas. “Estamos viviendo un momento histórico. Los datos que se generan a diario no tienen precedentes y sólo las máquinas que emulen el funcionamiento del cerebro humano nos pueden ayudar a aprovecharlos al máximo”, sostuvo Hanoi Morillo, jefe de rendimiento y transformación digital de Google.

3. Conclusiones

El machine learning o aprendizaje automático tiene especial efectividad en problemas de naturaleza compleja donde la aplicación de algoritmos ayuda a la obtención de soluciones precisas y con el consecuente ahorro de tiempo.

La capacidad para aprender es una de las características fundamentales de la inteligencia; el aprendizaje interviene en cualquier actividad que requiera inteligencia como por ejemplo el diagnóstico, la planificación, el lenguaje, las actividades sensor-motoras, etc. Por lo tanto, el aprendizaje automático juega un papel esencial en aplicaciones que han demostrado ser demasiado difíciles de programar manualmente, permitiendo construir algoritmos complejos que desarrollan la gran cantidad de aplicaciones actuales.

El éxito o fracaso de la aplicación de los métodos supervisados especialmente en los que se enfocan en las distancias depende de la técnica a aplicar y del conocimiento del experto para clasificar la clase.

La diversidad de métodos, técnicas y algoritmos, sean supervisados o no supervisados, enfocados a apoyar al aprendizaje automático o machine learning hace que esta área crezca a ritmo acelerado en la actualidad y sea el principal soporte de la inteligencia artificial al servicio del ser humano en todas las áreas, de ahí que radica su importancia en saberla aplicar correctamente.

4. Recomendaciones

Se recomienda que los temas investigativos en los centros de educación superior y los grupos de investigación utilicen este tipo de técnicas de aprendizaje automático, inteligencia artificial para apoyar sus procesos de desarrollo y solución de problemas del entorno.

En el sector educativo se hace necesario incrementar en los pénsum de estudios de ingeniería, temas como éstos, que permitan fortalecer los procesos de investigación y encontrar soluciones a diversos requerimientos de la sociedad, considerando que vivimos en una época en la que la generación de información se la hace a pasos agigantados.

La difusión de este tipo de técnicas de aprendizaje automático permite establecer mecanismos para que puedan proponer nuevos espacios de estudio, investigación e implementación a campos empresariales.

5. Bibliografía.

Alvarado-Pérez, J. C., Peluffo-Ordóñez, D. H., & Therón, R. (2015). Bridging the gap between human knowledge and machine learning. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 4(1), 54–64. <https://doi.org/10.14201/ADCAIJ2015415464>

Andrew, Y., Jordan, M. I., & Weiss, Y. (2001). On Spectral Clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems* (págs. 849-856). MIT Press.

Araujo, B. S. (2006). *Aprendizaje automático: conceptos básicos y avanzados: aspectos prácticos utilizando el software Weka*. Pearson Prentice Hall. Recuperado a partir de <https://books.google.com.ec/books?id=BCzUAQAACAAJ>

Basante, C., ortega, C., Peluffo, D., & Blanco, X. (2017). ESTUDIO COMPARATIVO DE TÉCNICAS SUPERVISADAS DE MACHINE LEARNING APLICADAS EN PROBLEMAS MÉDICOS. UTN. Recuperado a partir de https://issuu.com/utnuniversidad/docs/ebook_innovando_tecnologia_2017/91

Castro, C., Castro, A., Peluffo, D., & Castellanos, G. (2011). On the groups number estimation for unsupervised. *XV SIMPOSIO DE TRATAMIENTO DE SEÑALES, IMAGENES Y VISIÓN ARTIFICIAL - STSIVA 2011*.

Emtiyaz, S., & Keyvanpour, M. (2011). Customers Behavior Modeling by Semi-Supervised Learning in. *Advances in information Sciences and Service Sciences(AISS)*, 3(9). *Cornell University Library*.

Fontdeglòria, X. (2018, febrero 8). La policía china usa gafas con reconocimiento facial para identificar a sospechosos. *El País*. Recuperado a partir de https://elpais.com/internacional/2018/02/07/mundo_global/1518007737_209089.html

Hernández Bonilla, J. (2017, octubre 25). Así va la Inteligencia artificial y el aprendizaje automático de Google | ELESPECTADOR.COM. Recuperado a partir de <https://www.elespectador.com/tecnologia/asi-va-la-inteligencia-artificial-y-el-aprendizaje-automatico-de-google-articulo-719816>

IBM. (2012, junio 18). ¿Qué es Big Data? Recuperado el 19 de noviembre de 2017, a partir de <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>
Mitchell, T. M. (1997). *Machine Learning* (1a ed.). New York, NY, USA: McGraw-Hill, Inc.

Moya, R. (25 de marzo de 2016). *Jarroba*. Obtenido de <https://jarroba.com/que-es-el-clustering/>

Rodríguez Vázquez, S., Martínez Borges, A. V., & Lorenzo Ginori, J. V. (2015). Clasificación de células cervicales mediante el algoritmo KNN usando rasgos del núcleo, *10*(1), 15.
Sammur, C., & Webb, G. I. (2011). *Encyclopedia of Machine Learning*. Springer Science & Business Media.

Umaquina, A. C., Peluffo, D., Alvarado P., J. C., & Cabrera A., M. V. (2016). Estudio descriptivo de técnicas aplicadas en herramientas Open Source y comerciales para visualización de información de Big Data. En *Generando Ciencia: Memorias de las I Jornadas Internacionales de Investigación Científica UTN*. Recuperado a partir de https://www.researchgate.net/publication/312217978_Estudio_descriptivo_de_tecnicas_aplicad

Pasto, 19-20 de abril del 2018

Ana Umaquina, Luis Suárez, Omar Oña (Universidad Técnica del Norte / Ibarra-Ecuador)

as_en_herramientas_Open_Source_y_comerciales_para_visualizacion_de_informacion_de_Big
_Data