

# Aplicaciones de Machine Learning: Caso de Agrupación de datos mediante NLP con técnicas de aprendizaje supervisado

**Lady SANGACHA-TAPIA**

Unidad de Proyectos de Investigación I+D+i, Instituto Superior Tecnológico del Azuay con condición de Superior Universitario  
Cuenca, Ecuador  
lady.sangacha@tecazuay.edu.ec

**Yomar GONZÁLEZ-CAÑIZALEZ**

Facultad de Ingeniería Industrial, Universidad de Guayaquil  
Guayaquil, Ecuador  
yomar.gonzalezc@ug.edu.ec

**Rubén MANRIQUE-SUAREZ**

Facultad de Ingeniería Industrial, Universidad de Guayaquil  
Guayaquil, Ecuador  
ruben.manriques@ug.edu.ec

**Galo ESTUPIÑAN-VERA**

Facultad de Ingeniería Industrial, Universidad de Guayaquil  
Guayaquil, Ecuador  
galo.estupinanv@ug.edu.ec

**RESUMEN:** El trabajo desarrollado pretende identificar nuevas aplicaciones de conocimientos de machine learning como por ejemplo la ciencia e ingeniería de la computación para la industria, esto es posible agrupando datos bibliográficos donde los textos del lenguaje natural sean procesados a través del aprendizaje a la maquina de tipo supervisado, esta rama de la inteligencia artificial se agrupan los datos para clasificar el texto y finalmente etiquetarla en forma binaria, siendo aplicado en el caso de los algoritmos de optimización orientados a la investigación de operaciones. El método incluyó minería de datos bibliográficos de más de 1000 registros publicados en la base de datos relevantes como (Scopus, Web of Science, etc.), siguiendo la metodología SEMMA, para la identificación de patrones en los datos y predecir la tendencia en aplicaciones de Machine Learning en la industria con procesamiento de lenguaje natural, mediante un algoritmo de aprendizaje supervisado para la optimización del modelo de clasificación binaria por código con su lenguaje de programación de Python, utilizado en la ciencia de datos. La aplicación del modelo mostró un 99% asertivo en los resultados de confianza en el modelo de predicción.

**Palabras Claves:** Machine learning, algoritmos, procesamiento de lenguaje natural, tendencias de aprendizaje, industria 4.0, aprendizaje supervisado.

## ***Machine Learning Applications: Case of Data Clustering using NLP with Supervised Learning Techniques***

**ABSTRACT:** *The work developed aims to identify new applications of machine learning knowledge, such as computer science and engineering for industry, possible is by grouping bibliographic data where natural language texts must process through supervised machine learning, this branch of intelligence groups data data to classify the text and finally label it in binary*

*form, being applied in the case of optimization algorithms oriented to operations research. The method included bibliographic data mining of more than 1000 records published in the relevant database as (Scopus, Web of Science, etc.), following the methodology SEMMA, for identifying patterns in data and predicting the trend in Machine Learning applications in industry with natural language processing, using a supervised learning algorithm for the optimization of the binary classification model by code with its Python programming language, used in data science. The application of the model showed 99% assertive result of recall in the prediction model.*

**Keywords:** *Machine learning, algorithms, natural language processing, learning trends, industry 4.0, supervised learning.*

## **1. INTRODUCCIÓN**

Dentro del crecimiento tecnológico, se están produciendo constantemente nuevos descubrimientos en áreas y campos de conocimiento de la ciencia e ingeniería computacional como la inteligencia artificial y cada vez más amplios, donde se encuentran generando un impacto significativo en múltiples enfocado a la industria. Estos campos emergentes están impulsando la innovación y creando oportunidades sin precedentes como parte de las transformaciones educativas que servirán en vivir en la industria abriendo caminos que recorrer [1]. Los resultados de la investigación realizada hacen referencia a la aplicación de nuevos métodos bajo estudios utilizando procesamiento de lenguaje natural para predecir la tendencia en aplicaciones de Machine Learning en la industria para la educación que permitan proponer nuevas aplicaciones de aprendizajes interesantes, contribuyendo con los objetivos de la propuesta agenda para el 2030, el Desarrollo Sostenible Educación 2030. [2]

Por tal razón el banco mundial dentro de su informe llamado “Aprender a realizar la promesa de la educación” menciona que el 70% de los países del mundo están invirtiendo menos del 4% del PIB con referente a la educación, dando a conocer la importancia de los constantes cambios técnicos y tecnológicos [3] [4] para el funcionamiento de los aprendizajes dentro del complejo mundo de la industria, donde la contribución de los diferentes campos de conocimientos son perennes en la educación superior, esto se ve reflejado en los países del primer mundo como en los países de desarrollo que tienen su factor común que es la búsqueda continuo o metódica en las diferentes IES que esta sumergidos a los revolución de cambios por la construcción de nuevos conocimientos obligando a las nuevas evaluaciones [1]. La Organización para la Cooperación y el Desarrollo Económicos (OCDE, por sus siglas en inglés) utiliza esta medida, entre otras, para analizar y comparar diversos indicadores económicos y sociales entre los países miembros relacionados a la educación. El banco del desarrollo del mundo menciona que el obtener una educación universitaria por lo general implica adquirir una amplia gama de conocimientos y habilidades aplicables en diversos campos. Esto incluye habilidades de pensamiento crítico, resolución de problemas, comunicación y capacidad para trabajar en equipo. Estas competencias son muy buscadas por los empleadores, ya que contribuyen a la efectividad y productividad general de una persona en el lugar de trabajo, y que mejor en la aplicación de la industria que ofrece varios beneficios [5] como un aspecto más para la educación universitaria que desempeña un papel fundamental en la sociedad contemporánea, ya que busca alcanzar la calidad educativa necesaria para hacer frente a los desafíos planteados por la globalización. La docencia universitaria se enfrenta a la tarea de adaptarse a nuevos paradigmas con el fin de seguir siendo la principal fuente de transmisión del conocimiento a los estudiantes, así lo menciona uno de los países de Latinoamérica. Se encontró también la existencia de temas de tableros de mando para la toma de decisiones en la gestión administrativa de la educación superior la cual aportaría la alimentación de ello con la aplicación de nuevos algoritmos. [6]

Según en el informe del banco mundial expone que los cambios tecnológicos han llevado la transformación en el trabajo para una nueva era, una segunda era de las máquinas, la cuarta revolución industrial, por lo que muchos empleos desaparecerán, desvalorizando las habilidades humanas, pero por ahora los drásticos cambios pronosticados no se han extendidos en los países de altos ingresos y mucho menos en los ingresos bajos y medios [5]. Así mismo lo menciona la Unidad de Estudios del Empleo de la División de Desarrollo Económico de la Comisión Económica para América Latina y el Caribe (CEPAL), los grandes desafíos para nuevas oportunidades laborales debidos a los nuevos cambios tecnológicos [7]. En el informe de la Organización de las Naciones Unidas (ONU) del año 2023 en el enfoque de la educación menciona las nuevas tendencias tecnológicas que reemplaza los diferentes campos laborales por la segunda era tecnológica [8] [2], lo mismo establece en los informes 2022 de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), ofreciendo una estrategia de innovación tecnológica en la educación (2021-2025), donde contempla recomendaciones sobre el aprendizaje de la AI [9], guías y principios éticos del uso de la AI, las orientaciones para los educandos. [10]

Por lo concerniente los nuevos descubrimientos son cambios positivos en la formación como la ingeniería, eso se ha presentado antes donde el centro es el proceso de las

innovaciones [11] como en la actualidad que ya se presentan otros entornos de aprendizaje [12] y esto ya es posible con la aplicación de las nuevas tendencias tecnológicas como la aplicación de la inteligencia artificial (AI) donde el procesamiento de lenguaje natural (NLP) [13][14] como parte de la lingüística computacional [15] por ser una disciplina que combina la lingüística, ciencias de la computación y a AI, realiza búsquedas más complejas y precisas en repositorios y bases de datos académicos, donde las máquinas comprendan, interprete y generen lenguaje humano, aplicando sus técnicas en el procesamiento de lenguaje natural [16]. Aquí hay algunas formas en que el procesamiento natural de la lengua puede ser beneficioso como generación de resumen, clasificación de documentos, análisis de citas, extracción de información, todo referente a texto, se obtiene los indicadores de conocimiento para cubrir la enseñanza de estos cambios, el desafío que se avecinan con los nuevos campos de conocimiento, contribuyendo las nuevas líneas de investigación enfocados al caso de los algoritmos de optimización orientados a la investigación de operaciones como parte de la nueva era de las máquinas. Como anteriormente se indicó que la aportación de la aplicación de la inteligencia artificial, donde su abanico de diferentes sub-campos se plantea con la aplicación de la clasificación del procesamiento de lenguaje natural basado en machine learning [17], esto es posible con la aplicación de aprendizaje supervisado permitiendo la identificación de esos criterios por la minería de datos para el análisis de aproximadamente 1800 datos pero que aproximadamente 1000 datos fueron validados implementando la metodología SEMMA, permitiendo la predicción de los criterios de búsquedas que identifique la predicción de los nuevos campos de conocimiento contribuyendo a la agenda de educación del 2023. Consistió en la experimentación de clasificación de texto basado en machine learning [18] en 4 fases, la principal fue en recolectar artículos de alto impactos en bases científicas confiables como scopus, Web of site entre otros con la herramienta de un metadato, para ello era importante considerar los criterios como: que el artículo tenga código de ISBN, su presentación de DOI, en caso de presentar en revista, que se encuentre indexada por la web of Science, se puede verificar también en el elsevier de su página oficial de scopus, una vez que se verifico que producción o artículo científico era scopus se procedió en pre procesar los datos en conjunto con el análisis de sus variables, más adelante podrás conocer el detalle de cada fase, hasta eso puedes conocer que luego del pre procesamiento, la aplicación del método EDA en los datos que es el proceso de la exploración de datos, su aplicación de modelo, para finalmente evaluar su algoritmo permitió su predicción de indicadores. En ese proceso existe en aplicar la clasificación de texto [19][20] que hoy en día podemos utilizar la plataforma humans ai donde la idea nació de la aplicación del proceso de etiquetación, todo con el uso del lenguaje de alto nivel de Python. Los resultados de esta investigación se determino los indicadores utilizando la matriz de confusión donde su rendimiento de predicción, exactitud, el recall fue del 96% de asertividad.

## 2. REVISIÓN DE LITERATURA

### Machine Learning

En los años 50 nació el origen de aprendizaje de máquina, en ingles Machine learning aquella subrama de la inteligencia artificial que entrena aprendizajes a las máquinas para que piense, actúe y razone como los seres humanos de forma automatizada.

El origen del aprendizaje automático (machine learning) se remonta a varias décadas atrás, con contribuciones significativas de diferentes investigadores en el campo de la inteligencia artificial y la estadística. No hay un único origen o autor definitivo, ya que el aprendizaje automático ha evolucionado a lo largo del tiempo gracias a numerosos avances teóricos y prácticos. Se menciona que las bases teóricas para la idea de una máquina universal desarrollada por Turing. Esta noción se considera un precursor crucial para el desarrollo de la computación y el aprendizaje automático. "Podemos ver claramente que no hay límite teórico a lo que una máquina universal de Turing puede hacer de manera completamente automática", afirmó Turing [21]. Sin embargo, el trabajo pionero del científico Arthur Samuel en la década de 1950 marca un momento crucial en el desarrollo del aprendizaje automático. Samuel utilizó el término "aprendizaje automático" para describir cómo las computadoras pueden aprender y mejorar automáticamente de los datos sin estar programadas explícitamente para hacerlo. Su objetivo era crear algoritmos que pudieran jugar al ajedrez de manera autónoma y mejorar su desempeño a medida que acumulaban experiencia. [22]

### Tipos de Aprendizajes

Cada tipo de aprendizaje tiene sus enfoques y características específicos, donde se le entrega a la máquina para que aprenda, y esto se ha ido llevando en los diferentes campos de la ciencia que llevan a la automatización de tareas que podrían requerir de una intervención humana, además de la toma de decisiones basadas en datos, personalización, detección de anomalías en fraudes [23], avances médicos [24] entre otros, un potencial de una amplia gama de industrias y aplicaciones. Existían 3 tipos de aprendizaje el Aprendizaje supervisado, Aprendizaje no Supervisado y Refuerzo según la naturaleza de los datos que son recibidos. En el transcurso del tiempo se ha ido evolucionando a un nuevo aprendizaje que es semi supervisados y el aprendizaje activo, aunque aun se ha encontrado que Bernard Schölkopf: ha realizado importantes investigaciones en el campo del aprendizaje se-mis supervisado junto con sus colaboradores. Su trabajo en métodos de aprendizaje basados en el núcleo ha tenido un impacto en la creación de algoritmos semi supervisados [25]. La aplicación de los diferentes algoritmos dependerá de la situación de la solución del problema. Se menciona que los años de 1805 fue publicado la regresión por el método de mínimos cuadrados y el año de 1809 los autores Legendre y Gauss aplicaron unos métodos a través de la observación astronómicas, los cuerpos celestes para la determinación del problema [26]. Varios investigadores y matemáticos han contribuido al desarrollo de las técnicas de regresión. Para predecir un valor continuo o numérico, se utiliza la regresión. El objetivo de un algoritmo de regresión es descubrir una función que conecte las variables de entrada una variable de salida continua. Por ejemplo, la regresión se puede utilizar para predecir el precio de una casa en función de su ubicación, tamaño, número de habitaciones y otros factores. La regresión lineal, la regresión logística, la regresión polinómica y la regresión de vectores de soporte son algunos ejemplos comunes de algoritmos de regresión. A través de la ecuación:

$$y = w x + b \quad (1)$$

Donde  $w$  representa la pendiente (inclinación) de la línea recta y  $b$  es la intersección con el eje  $y$ . Cuando se intenta predecir una categoría o clase específica, se utiliza la clasificación [20]. Un algoritmo de clasificación tiene como objetivo asignar objetos de entrada en una clase o categoría predeterminada en función de

sus características. Por ejemplo, basándose en el contenido y las características de un correo electrónico, se puede usar la clasificación para determinar si es spam o no. Los árboles de decisión, los bosques aleatorios, las redes neuronales, las máquinas de vectores de soporte (SVM), clasificación binaria y la clasificación bayesiana son algunos ejemplos comunes de algoritmos de clasificación.

### Pasos para entrenar un modelo en NLP

Para ello existe una serie de pasos para entrenar un modelo, un proceso utilizado para realizar las predicciones con la aplicación de los algoritmos. Donde los datos recibidos y patrones son relacionados para entrenarlos, obteniendo nuevos datos, esta herramienta de uso matemática o algorítmica se puede aplicar para que la máquina aprenda y tomar decisiones a través de la información, para ello machine learning nos da pautas para entrenar un modelo con la finalidad que la máquina aprenda, una guía para aplicar en cualquier modelo supervisado que se menciona a continuación:

- a) Tener claro el problema para correlacionar el objetivo a tratar.
- b) Colección o recolección de datos, una parte que se podría decir es donde se lleva la mayor cantidad de tiempo.
- c) Pre procesamiento y procesamiento de datos, con la ayuda de EDA.
- d) Selección del modelo para la aplicación del algoritmo
- e) Evaluación del modelo para conocer el porcentaje de asertividad o predicción.

### Procesamiento de Lenguaje Natural

Estos representan cada palabra mediante embeddings o vectores numéricos, que codifican su significado y función en las oraciones. Actores que integran dentro del procesamiento de Lenguaje Natural. Se conoce que existe una intervención de machine learning, Deep learning, procesamiento de lenguaje natural sin embargo en lo que se diferencia es la forma de interpretar el lenguaje natural, es decir el idioma lingüístico, por lo tanto, los actores que intervienen en NLP es:

- a) NLU Natural language understanding
- b) NLG Natural language generation
- c) TM Text mining

El TM es una minería de datos, descubre información asociada que se aplica en el texto que permite estructurar y transformar el texto. NLU encargada de entender el texto, una comprensión lingüística aplicado en las traducciones automática, las preguntas y respuestas, o las categorizaciones de textos sin olvidar el análisis de sentimiento. NLG encargada de la generación del lenguaje, ayudando a las creaciones de interfaces que realizan conversaciones. En la cual posee sus técnicas de procesamiento de lenguaje natural en la inteligencia artificial para texto [16].

### 3. APLICANDO LA METODOLOGÍA

La aplicación de herramientas de enseñanza esta contribuyendo significativamente, con gran uso en la medicina [27] sin embargo ya se esta extendiendo la aplicación en otros campos, que en nuestro caso es identificar nuevas aplicaciones de conocimientos de machine learning en la enseñanza para la industria, con la minería de datos de proyectos con la extracción de datos bibliográficos que serán transformados permitiendo la

identificación de los patrones y predicción de la tendencia en aplicaciones utilizando la metodología SEMMA.

Debido a su generación de conocimientos, se consideró la actividad sistemática de investigación básica, ya que cumple la función social que vincula con la sociedad, en este caso al enfoque de educación superior por medio de identificar los criterios para la búsqueda de información, considerando la metodología cuantitativa. Se llevo 4 momentos considerados como fases, para llevar a cabo el diseño del modelo que se planteó en el objetivo descrita en el resumen, donde se basó en la metodología de Sample Explore Modify Model Assess (SEMMA), de las siguientes formas:

Fase 1- Identificación del problema, esto permitió detectar la solución necesaria para alcanzar el objetivo e identificar cual sería la muestra para tomar, necesaria a pensar que modelo de aprendizaje automático se pudo optar, que sea útil en la identificación de criterios en la industria 4.0 para las líneas de investigación basado en machine learning con el procesamiento de texto (NLP). Como anteriormente se indico la revolución de la industria, obliga a que los paradigmas de la educación tengan que alinear-se a la necesidad de que se viene donde las nuevas líneas de investigación son necesarias por los cambios tecnológicos en las diversas disciplinas de conocimientos, por lo que esto nos llevo a pensar ¿Qué modelo?, ¿Qué tipo de aprendizaje aplicar?, ¿Qué algoritmo aplicar? y finalmente ¿El modelo arquitectónico con aprendizaje supervisado del procesamiento de lenguaje natural basado en machine learning, permitirá la identificación de herramienta a la aplicación de conocimiento de machine learning?.

Fase 2 – Implica la colección de datos, para la llamada Data set, en esta fase se recolectó aproximadamente mas de 1800 artículos científicos de alto impacto desde varias fuentes confiables en dos mecanismos:

- a) El uso de los gestores bibliográficos
- b) Descargas de artículos científicos en revistas nacionales e internacionales, incluso de publicaciones de eventos científicos en scopus.

Cualquiera de los dos mecanismos pasa a un proceso de validación con la intención de verificar si son de scopus y una forma es realizarlo por medio de la página scimago o scopus, de lo que se ha probado con las distintas palabras y jugar con ellas [33].

Esos datos son colocados en una matriz de meta análisis con un tipo de archivo de .csv, para ello ya se había analizado con criterios bibliométricos, los campos requeridos fue el nombre del artículo, abstract, año de publicación, código ISSN o ISBN, si es articulo etc. Hasta ahí se podría decir que tenemos la data preparada que contiene contenidos de conocimientos necesaria con contenidos que aplican en la industria como las automatizaciones, aplicación de machine learning, algoritmos, etc.

Esta captación de artículos que ya habíamos mencionado que son de gran impacto fueron colocados dentro de los gestores bibliográficos, con la finalidad de automatizar el proceso de validación.

Fase 3 – Preparación de datos para su manipulación de datos antes de darle un entrenamiento a la maquina, debe ser pre procesadas, sin duda un trabajo de hormiga en la fase 2, que una

vez terminada se logró recolectar aproximadamente 1800 registros que correspondía de artículos científicos, nos tomo 7 meses en solo recolectar sin embargo después de revisarla y verificar fue-ron 1064 registros que cumplían de ser producciones de alto impacto relacionadas a la industria de 4.0 con algoritmos de investigación de operaciones, algoritmos de optimización y sus aplicaciones, para esta fase se eliminaron los que no cumplían con el parámetro de criterios bibliométricos, además de campos innecesarios que estaban incompletos. Una vez listo se procede a cargar los datos de la llamada Data set, utilizamos el lenguaje de Python con una herramienta científica de datos para el debido proceso de exploración, preprocesamiento y procesamiento.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

**Figura 1** Librerías de Python a utilizar para manipulación de datos para la utilización de código y correr el proceso de exploración. 2023.

En la Fig. 1 nos muestra una parte de la data con el uso de comando Python, un lenguaje que ayuda a proporcionar una inspección directa e indirecta de datos, para la aplicación exploración de análisis de datos (EDA), el inicio de encontrar correlación de datos para aplicar el pre-procesamiento. Por medio de la exploración se identifico los valores isnull que comprende de valores vacíos o nulos, eliminando aquellos valores. Para este proceso teníamos claros las variables independientes y dependiente, poniendo en consideración el modelo algorítmico para la clasificación de texto y la columna que nos permitía aplicarlo en el resumen o abstract que nos proporcionaba nuestro meta análisis. El mismo que se obtuvo los datos pre entrenados como año de publicación del producto, fuente o link de acceso o doi, tipo de producto (hace referencia si es articulo, libro, etc), autores, apunte (Breve tip del documento), cita apa 7 (Texto original,citado), referencia en apa 7, Fecha de Publicacion dd/mm/aaaa, ISBN/ISSN, Titulo del Artículo, Nombre de la revista o fuente donde fue publicado, Resumen que trae el Artículo, Hallazgos del Artículo usando palabras, indicadores o variables destacadas. Evite redacciones vacías o generalizaciones, Abstract traducido a espaniol usando alguna herramienta, Conclusiones que trae el Artículo en español.

Después de varias experimentaciones e investigaciones descubrimos que se podría clasificar de dos formas tener una conversación del experto que nos proporcione la lista de clasificación de texto orientadas a machine learning en la industria o la utilización de la herramienta de inteligencia artificial que nos entregue la clasificación de texto de forma automatizadas, se realizo de las 2 formas. Lo importante de esta fase que una vez que se clasifico, manualmente se etiqueto donde se coloco con 1 todos aquellos artículos que cumplían en cada una de las 22 columnas de la clasificación y 0 los que no. Este proceso se aplico con la herramienta científica de datos.

Fase 4 – Luego del pre-procesamiento y procesamiento se procede a modelar el algoritmo de clasificación binaria y red neuronal. Como primera instancia se elimina las columnas no necesarias, luego revisamos para verificar el tratamiento de los datos nulos como se observa la figura 4. Al realizar el proceso quedaron 1058 registros y 22 columnas. Luego de ello se crea un

nuevo archivo de datos procesados para la aplicación binaria con sus librerías de keras, con esto se entrena la variable a utilizar dividiendo los inputs y los outputs, eso facilita en el testeó, prediciendo con un algoritmo de regresión, completando el entrenamiento. La utilización de las fórmulas matemáticas es indispensable, se utilizará la de regresiones al modelo binario.

$$Y=b_0+b_1*X_1 \tag{2}$$

$$Y=b_0+b_1*X_1+b_2*X_2+\dots+b_n*X_n \tag{3}$$

Con la intención de utilizar una vez detectado la variable independiente.

Fase 5- Esta fase es la valoración o evaluación, en la selección de entradas porque el modelo desconoce y procede en la verificación de presión, alcanzando el 99% de pre-dicción utilizando las variables de machine learning y modelos predictivos. Para visualizar esa evaluación se aplico la matriz confusión que nos ayuda evaluar el rendimiento del modelo de clasificación por cada categoría como se puede apreciar en la Figura 6. Además, se aplico la precisión recall, esto ayudó a observar las instancias relevantes. En esta fase se identifica esas variables de entradas independiente para obtener nuestra dependiente.

```
acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print("accuracy_score : ", acc)
print("precision_score : ", prec)
print("recall_score : ", recall)

accuracy_score : 0.9952830188679245
precision_score : 1.0
recall_score : 0.9896907216494846

Crosstab devuelve la tabla de contingencia resultante de cruzar dos o más campos de un dataframe.

titanic_crosstab

y_test 0.0 1.0
y_pred
0.0    115  1
1.0     0  96
```

Figura 2 Resultados de predicción

Se pudo aplicar otras experimentaciones algorítmicas con otros modelos como multiclase y multietiqueta pero se obtuvieron mejores resultados, más bien se aprecio mejores resultados en los modelos de algoritmo binario y de redes neuronales.

De igual forma los modelos aplicados ha sido de aprendizaje supervisado.

Podemos observar la figura 3 el mismo mecanismo que se utilizo, pero con el algoritmo de red neuronal, teniendo una prueba de accuracy del 0.69, donde en la grafica de la figura da a tendencia positiva de buena curva roc.

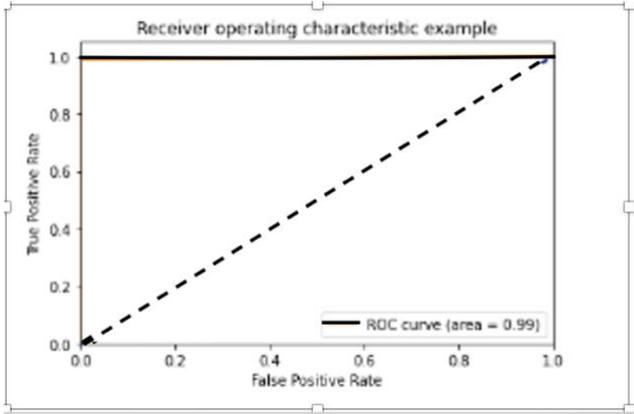


Figura 3 Resultado aplicando algoritmo de clasificación binaria. 2023

La aplicación de herramientas científicas con el lenguaje de Python nos señala que tiene tendencia a 1 con el uso del dominio de conocimiento de algoritmo predictivo. Con la ayuda del lenguaje de Python y la herramienta para tratar de visualizar e interpretar los datos que se ha recibido. Podemos apreciar mas resultados en la tabla 1 donde nos muestra la calidad del algoritmo. Se puede conocer términos que mas que nuevos son palabras de inglés. A continuación, nos da a conocer los resultados obtenidos una vez realizado la arquitectura del modelo creado o la aplicación de algoritmos en el modelo.

| Descripción       | Algoritmo de clasificación Binaria | Algoritmo de red neuronal |
|-------------------|------------------------------------|---------------------------|
| accuracy_score :  | 0.99                               | 0.69                      |
| precision_score : | 1.00                               | 1.00                      |
| recall_score :    | 0.98                               | 1                         |

Tabla 1 Tabla comparativa para visualizar la calidad de los algoritmos de clasificación binaria y una red neuronal.

Una vez obtenida la arquitectura, se puede ir experimentando con las diferentes variables de la data set obtenida, incluso se puede agregar más técnicas de pre procesamiento pero en texto. Ese proceso contempla como la ultima fase en la aplicación del algoritmo, que no se contemplo en el resumen.

4. RESULTADOS

El desarrollo bibliométrico permitió identificar las nuevas aplicaciones de conocimientos de machine learning en la enseñanza que permite abordar el observatorio de creación de nuevas líneas de investigación para la industria obteniendo un nuevo conocimiento descubriendo cuando se realiza el análisis de contenidos, aplicado con Python con herramienta tecnológica, aplicando el tipo supervisado con el algoritmo de clasificación de procesamiento de lenguaje natural, dando el 99% de predicción que el conocimiento es “machine learning y modelos predictivos”, utilizando aplicaciones informáticas sin embargo realizando con la aplicación del algoritmo de redes neuronal nos vota el 69%, aplicando 16 capas de entrada, podemos concluir que a mayor capas mejores resultados.

5. REFERENCES

- [1] G. S. Bueno, “Innovación en la Educación Superior : Paradigmas , Transformaciones y Innovation In Higher Education : Paradigms , Transformations And Good Educational Practices,” pp. 6708–6723, 2023.
- [2] UNESCO, “Transformar la educación para el futuro,” *UNESCO DOC Bibl. Digit.*, p. 16, 2022.
- [3] J. Southworth *et al.*, “Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy,” *Comput. Educ. Artif. Intell.*, vol. 4, no. January, p. 100127, 2023.
- [4] J. P. Aguayo Litardo, M. S. Paredes Rodríguez, and D. J. Paredes Rodríguez, “Innovación educativa: identificación de aspectos relevantes en los últimos años,” *Cienc. Lat. Rev. Científica Multidiscip.*, vol. 7, no. 1, pp. 10805–10817, 2023.
- [5] World Development Report, “Overview: Learning to realize education’s promise,” *World Development Report 2018: Learning to Realize Education’s Promise*. pp. 1–35, 2017.
- [6] E. A. Varela, I. L. Acosta, M. C. Toapanta, M. A. Sanchez, S. M. Medina, and F. M. Malla, “Proposal of scorecards for decision making in the administrative management of Higher Education | Propuesta de Tableros de Mando para la Toma de Decisiones en la Gestión Administrativa de la Educación Superior,” in *CICIC 2019 - Novena Conferencia Iberoamericana de Complejidad, Informatica y Cibernetica, Memorias*, 2019, vol. 2, pp. 32–37.
- [7] J. Weller, S. Gontero, and S. Campbell, “Cambio tecnológico y empleo : una perspectiva latinoamericana,” *Macroecon. del Desarro.*, vol. 201, no. 1680–8851, pp. 13–17, 2019.
- [8] P. C. Maldonado, “DOCUMENTOS DE PROYECTO Caracterización de la demanda laboral en el Ecuador con información administrativa,” 2019.
- [9] X. Wang *et al.*, “What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis,” *Comput. Educ.*, vol. 194, p. 104703, Mar. 2023.
- [10] OEI, *El Futuro de la Inteligencia Artificial en educación en América Latina*, vol. 85. 2023.
- [11] V.-G. Luis Roberto, “La educación en ingeniería en el contexto global: propuesta para la formación de ingenieros en el primer cuarto del Siglo XXI,” *Ing. Investig. y Tecnol.*, vol. 14, no. 2, pp. 177–190, 2013.
- [12] J. Gómez Amaya, D. A. Bautista Díaz, and J. F. Cortés Garnica, “SCAECH: herramienta basada en inteligencia artificial para la evaluación del aprendizaje en entornos constructoristas,” *Rev. Educ. en Ing.*, vol. 18, no. 35, pp. 1–9, 2023.
- [13] L. Sans, I. Vallvé, J. Teixidó, J. M. Picas, J. Martínez-Roldán, and J. Pascual, “La era del big data: análisis del lenguaje natural mediante la aplicación de folksonomía,” *Nefrología*, vol. 42, no. 6, pp. 680–687, 2022.
- [14] D. Dessí, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, and E. Motta, “SCICERO: A deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain,” *Knowledge-Based Syst.*, vol. 258, p. 109945, 2022.
- [15] J. Benchimol, S. Kazinnik, and Y. Saadon, “Text mining methodologies with R: An application to central bank texts,” *Mach. Learn. with Appl.*, vol. 8, no. March, p. 100286, 2022.
- [16] R. J. Celi-Parraga, E. A. Varela-Tapia, I. L. Acosta-Guzmán, and N. R. Montañó-Pulzara, “Técnicas de procesamiento de lenguaje natural en la inteligencia artificial conversacional textual,” *AlfaPublicaciones*, vol. 3, no. 4.1, pp. 40–52, 2021.
- [17] M. S. Rahman, T. Ghosh, N. F. Aurna, M. S. Kaiser, M. Anannya, and A. S. M. S. Hosen, “Machine learning and internet of things in industry 4.0: A review,” *Meas. Sensors*, vol. 28, no. June, p. 100822, 2023.
- [18] C. Aracena, F. Villena, F. Arias, and J. Dunstan, “Aplicaciones de aprendizaje automático en salud Applications of machine learning in healthcare,” *Rev. Clínica Las Condes*, vol. 33, no. 6, pp. 568–575, 2022.
- [19] A. Fern *et al.*, *Ecuador: realidad nacional*. 2020.
- [20] J. M. M.-G. Paula Martín-Climent, “Aplicación de la inteligencia artificial en el laboratorio de reproducción asistida,” *Angew. Chemie Int. Ed.* 6(11), 951–952., vol. 3, no. 1, pp. 10–27, 2022.
- [21] C. F. Huws and J. C. Finnis, “On computable numbers with an application to the Alan Turing problem,” *Artif. Intell. Law*, vol. 25, no. 2, pp. 181–203, 2017.
- [22] E. Alcobilla Ferrara, “Introducción al machine learning en Senología,” *Rev. Senol. y Patol. Mamar.*, vol. 36, no. 4, p. 100503, Oct. 2023.
- [23] P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, “Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques,” *Procedia Comput. Sci.*, vol. 218, pp. 2575–2584, 2023.
- [24] E. Basáez and J. Mora, “Salud e inteligencia artificial : ¿ cómo hemos evolucionado ? Artificial intelligence in health : where are we in 2022 ?,” *Rev. Clínica Las Condes*, vol. 33, no. 6, pp. 556–561, 2022.
- [25] G. Valle, “Integración de rasgos y aprendizaje semi-supervisado para la clasificación funcional de enzimas utilizando K-medias de Spark,” vol. 14, no. 4, pp. 134–161, 2020.
- [26] J.-J. Samueli, “Legendre et la méthode des moindres carrés,” *BibNum*, pp. 0–15, 2010.
- [27] J. M. Rosa and E. L. Frutos, “Health data science: Challenges and opportunities in Latin America,” *Rev. Medica Clin. Las Condes*, vol. 33, no. 6, pp. 591–597, 2022.