

**UNIVERSIDAD MAYOR DE SAN ANDRÉS**  
**FACULTAD DE CIENCIAS PURAS Y NATURALES**  
**CARRERA DE INFORMÁTICA**



**INTELIGENCIA ARTIFICIAL**  
**“CLASIFICACIÓN AUTOMÁTICA DE NOTICIAS BOLIVIANAS**  
**MEDIANTE APRENDIZAJE AUTOMÁTICO”**

ESTA INVESTIGACIÓN ES CON FINES EDUCATIVOS, PARA DEMOSTRAR LO  
MUCHO QUE PUEDE REALIZAR UNA INTELIGENCIA ARTIFICIAL

**POR: YOEL TICONA LAURA**  
**EVER EMERSON JURADO**  
**EDWIN ALIAGA YUJRA**  
**TUTOR: CHAT GPT, GEMINI**  
**NUESTRA SEÑORA DE LA PAZ - BOLIVIA**

**2025**

## LICENCIA DE USO

Yo, Yoel Ticona, Ever Jurado y Edwin Aliaga titulares del trabajo de titulación denominado “*Clasificación automática de noticias bolivianas mediante aprendizaje automático*”, presentado como requisito para la obtención del título de Licenciatura en Informática, autorizo a la Universidad [Nombre de tu Universidad] a archivar y publicar este documento en su repositorio institucional, con fines académicos y de investigación.

Declaro que este trabajo es de nuestra autoría, que las fuentes utilizadas han sido debidamente citadas y que no infringe derechos de propiedad intelectual. La presente obra puede ser consultada y reproducida parcialmente, citando la fuente, siempre que no sea con fines comerciales ni se realicen modificaciones sustanciales al contenido original.

La obra se distribuye bajo los términos de la licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional.

Toda la tesis realizada con IA fue realizada con hechos académicos sin desprestigiar ningún trabajo ni plagios, ni dar la idea de la facilidad de realizar una tesis.

## AGRADECIMIENTOS

A Dios, por darme la vida, la fuerza y la perseverancia para culminar esta etapa tan importante.

A mi familia, por su apoyo incondicional, su paciencia y por creer en mí aún en los momentos más difíciles.

A mi tutor de tesis, **[Nombre del Tutor]**, por su guía, experiencia y compromiso durante el desarrollo de esta investigación.

A mis docentes y compañeros de la carrera de Informática, quienes aportaron a mi formación con sus conocimientos y motivación.

A la Universidad Mayor de San Andrés, por brindarme la oportunidad de formarme como profesional y por ser el espacio donde nacieron muchas de mis ideas y aspiraciones.

A todos quienes de alguna forma aportaron a que este trabajo sea una realidad, gracias.

***Plantilla generada con inteligencia Artificial***

## RESUMEN

Esta investigación presenta una propuesta para la clasificación automática de noticias bolivianas utilizando técnicas de aprendizaje automático y procesamiento de lenguaje natural (PLN). Se identificó la necesidad de categorizar noticias digitales de manera eficiente y adaptada al contexto lingüístico del español boliviano. El estudio se enmarca en una metodología cuantitativa y experimental, donde se analizaron algoritmos como Naive Bayes, Máquinas de Soporte Vectorial (SVM) y redes neuronales. Se construyó un corpus representativo y se aplicaron técnicas de preprocesamiento textual. Los resultados teóricos sugieren que la implementación de estos modelos permite clasificar noticias en categorías temáticas con altos niveles de precisión. La investigación busca contribuir al desarrollo de herramientas inteligentes para el análisis de medios bolivianos, y plantea una base sólida para futuras aplicaciones tecnológicas en el ámbito del periodismo digital.

**Palabras clave:** clasificación automática, noticias digitales, aprendizaje automático, PLN, español boliviano.

## ABSTRACT

This research presents a proposal for the automatic classification of Bolivian news using machine learning techniques and natural language processing (NLP). The need to efficiently categorize digital news in a way that is linguistically adapted to Bolivian Spanish was identified. The study follows a quantitative and experimental methodology, analyzing algorithms such as Naive Bayes, Support Vector Machines (SVM), and neural networks. A representative news corpus was built, and standard text preprocessing techniques were applied. Theoretical results suggest that these models are capable of categorizing news into thematic areas with high accuracy. This research aims to contribute to the development of intelligent tools for analyzing Bolivian media and provides a foundation for future technological applications in digital journalism.

**Keywords:** automatic classification, digital news, machine learning, NLP, Bolivian Spanish.

## ÍNDICE

CAPÍTULO I.....	1
INTRODUCCIÓN.....	1
1.1    ANTECEDENTES.....	2
1.1.1    Antecedentes institucionales .....	2
1.1.2    Proyectos similares.....	2
1.2    OBJETIVO DE ESTUDIO O INVESTIGACIÓN .....	4
1.3    PLANTEAMIENTO DEL PROBLEMA.....	4
1.4    HIPOTESIS .....	5
1.4.1    Hipótesis General: .....	5
1.4.2    Hipótesis Específicas: .....	5
1.5    JUSTIFICACIÓN.....	7
1.6    OBJETIVOS.....	8
1.6.1    Objetivo general .....	8
1.6.2    Objetivos específicos.....	8
1.7    ALCANCE Y LÍMITES.....	8
1.7.1    Alcances .....	8
1.7.2    Límites.....	8
1.8    IMPORTANCIA DEL ESTUDIO .....	9
CAPÍTULO II.....	10
MARCO TEÓRICO .....	10
2.1.    Inteligencia Artificial y Aprendizaje Automático.....	10
2.2.    Procesamiento de Lenguaje Natural (PLN) .....	11
2.3.    Clasificación de Textos .....	11
2.4.    Clasificación de Noticias Digitales .....	13
2.5.    Estado del arte .....	13
CAPÍTULO III .....	15

MARCO METODOLÓGICO .....	15
3.1. Tipo de investigación .....	15
3.2. Enfoque de la investigación .....	15
3.3. Diseño de la investigación.....	15
3.4. Población y muestra .....	16
3.5. Técnicas e instrumentos de recolección de datos.....	17
3.6. Análisis de datos.....	17
3.7. Cronograma de actividades .....	17
3.8. Operacionalización de Variables .....	17
<b>Variable 1: Algoritmos de Clasificación</b> .....	17
<b>Variable 2: Precisión de la Clasificación</b> .....	18
<b>Variable 3: Adaptación del PLN al Español Boliviano</b> .....	18
CAPÍTULO IV .....	19
MARCO TRIBUNAL .....	19
4.1. Protección de Datos Personales.....	19
4.2. Uso Ético de la Inteligencia Artificial .....	19
4.3. Cumplimiento de la Normativa Internacional.....	19
4.4. Consentimiento Informado en la Recolección de Datos .....	19
CAPÍTULO V.....	20
CONCLUSIONES Y RECOMENDACIONES .....	20
5.1. CONCLUSIONES .....	20
5.2. RECOMENDACIONES .....	20
BIBLIOGRAFÍA .....	22

## ÍNDICE DE FIGURAS

Ilustración 1. Ejemplo de proceso de aprendizaje automático. ....	10
Ilustración 2. Diagrama de técnicas de preprocesamiento de texto en PNL. ....	11

Ilustración 3. Diagrama de ventajas y desventajas de los algoritmos de clasificación en Machine Learning .....	12
--	----

### **ÍNDICE DE TABLAS**

Tabla 1 Cuadro Comparativo: Objetivos, Problemas e Hipótesis .....	5
Tabla 2. Tabla comparativa de Trabajos Relacionados.....	14



# CAPÍTULO I

## INTRODUCCIÓN

En la era de la información, el acceso a las noticias se ha vuelto instantáneo, diverso y, muchas veces, abrumador. Bolivia no es ajena a este fenómeno, ya que cada día se generan numerosos contenidos informativos a través de medios digitales, redes sociales y portales de noticias locales. Esta sobreabundancia informativa, conocida como infoxicación, plantea el desafío de organizar y filtrar grandes volúmenes de datos para que los usuarios puedan acceder a información relevante de manera eficiente.

En este contexto, el uso de técnicas de **aprendizaje automático (machine learning)** ha cobrado importancia como una solución innovadora para clasificar, filtrar y analizar grandes cantidades de textos de forma automatizada. Estas técnicas permiten entrenar modelos capaces de identificar patrones y categorías en textos, como noticias, sin intervención manual directa, mejorando así los procesos de organización y recuperación de la información (Alpaydin, 2020).

La clasificación automática de noticias consiste en asignar etiquetas temáticas —por ejemplo, política, economía, deportes, salud, entre otras— a textos periodísticos en base a su contenido. En países como Bolivia, donde los medios de comunicación digitales han crecido considerablemente en la última década, esta herramienta podría facilitar la estructuración de bases de datos noticiosas y mejorar la experiencia de los usuarios que consumen información diariamente.

Además, este tipo de clasificación automatizada tiene múltiples aplicaciones, desde la creación de resúmenes personalizados hasta la detección de noticias falsas, temas sensibles o tendencias emergentes. En el ámbito académico y de investigación, representa una oportunidad para combinar disciplinas como el procesamiento de lenguaje natural (PLN), la minería de datos y la inteligencia artificial, en beneficio de la sociedad y el acceso responsable a la información (Manning, Raghavan & Schütze, 2008).

Esta investigación se propone estudiar y aplicar modelos de aprendizaje automático para la **clasificación automática de noticias generadas en Bolivia**, evaluando su eficacia en el contexto local y utilizando corpus textuales representativos. Se espera que este trabajo no solo contribuya al desarrollo de soluciones tecnológicas aplicables al entorno periodístico, sino que también abra nuevas líneas de investigación en el área del procesamiento de texto en español y, específicamente, en español boliviano.

## 1.1 ANTECEDENTES

### 1.1.1 Antecedentes institucionales

En el contexto boliviano, el uso de herramientas de inteligencia artificial y aprendizaje automático en medios de comunicación aún se encuentra en una etapa incipiente. Aunque algunos medios digitales, como *La Razón*, *El Deber* o *Página Siete*, han desarrollado plataformas web con sistemas de búsqueda y categorización básica, estos procesos suelen ser manuales o semiautomáticos, sin el uso explícito de modelos de clasificación automática basados en aprendizaje de máquina.

A nivel académico, la Universidad Mayor de San Andrés (UMSA), a través de la carrera de Informática y otros programas afines, ha impulsado investigaciones relacionadas con minería de datos, procesamiento de lenguaje natural y desarrollo de sistemas inteligentes. Sin embargo, hasta la fecha no se ha encontrado documentación formal que indique la existencia de proyectos institucionales centrados específicamente en la clasificación automática de noticias nacionales mediante algoritmos de aprendizaje automático. Esto evidencia un espacio aún poco explorado, pero con alto potencial investigativo y de aplicación práctica.

Cabe mencionar que algunas tesis de grado han abordado temáticas cercanas, como análisis de sentimiento en redes sociales o sistemas de recomendación en el ámbito educativo, lo cual muestra una creciente apertura hacia las tecnologías emergentes en el entorno académico boliviano.

### 1.1.2 Proyectos similares

A nivel internacional, varios estudios han demostrado la eficacia de los modelos de aprendizaje automático en la clasificación de noticias. Por ejemplo, el uso de algoritmos como Naive Bayes, Support Vector Machines (SVM) y redes neuronales ha sido ampliamente validado en tareas de categorización automática de textos (Sebastiani, 2002; Zhang et al., 2015). Estos modelos han sido implementados en grandes plataformas de contenido como Google News o Yahoo News, permitiendo organizar miles de artículos diarios por categorías temáticas.

En América Latina, iniciativas como *LATAM NLP* han promovido el uso de modelos de PLN entrenados específicamente en español y portugués, mejorando los resultados en tareas como clasificación de textos y detección de tópicos. En países como México, Colombia y Argentina, se han desarrollado sistemas automatizados de análisis de medios para identificar discursos políticos, desinformación y tendencias mediáticas.

En Bolivia, aunque no existen muchos estudios formales publicados, algunos desarrolladores independientes y medios digitales han comenzado a experimentar con tecnologías como el análisis semántico o la búsqueda inteligente, lo cual indica un creciente interés por la automatización de procesos editoriales. Sin embargo, la ausencia de modelos entrenados con corpus locales en español boliviano representa una limitación importante, lo que refuerza la pertinencia de esta investigación.

En el ámbito boliviano, aún no se han identificado tesis que aborden de manera directa la clasificación automática de noticias utilizando técnicas de aprendizaje automático. Sin embargo, existen investigaciones académicas que han explorado temas cercanos, especialmente dentro del área del procesamiento de lenguaje natural (PLN) y la minería de datos, las cuales resultan relevantes por las metodologías empleadas y su posible adaptación al presente estudio.

- Título: MODELO DE ANÁLISIS DE NORMAS DE GOBIERNOS AUTÓNOMOS BASADO EN APRENDIZAJE AUTOMÁTICO

*Autor:* No especificado

*Institución:* Universidad Mayor de San Andrés (UMSA)

*Año:* No especificado

*Descripción:* Esta tesis investiga el uso de técnicas de aprendizaje automático para facilitar la detección de temas en una base de datos que contiene normas de gobiernos autónomos en Bolivia desde 2010 hasta 2019. Aunque no se centra en noticias, utiliza aprendizaje automático para clasificar documentos textuales, lo que puede ofrecer metodologías aplicables a la clasificación de noticias.

- Título: TRADUCTOR AUTOMÁTICO ENTRE ESPAÑOL Y QUECHUA BASADO EN PROCESAMIENTO DE LENGUAJE NATURAL

*Autor:* No especificado

*Institución:* Universidad Mayor de San Andrés (UMSA)

*Año:* No especificado

*Descripción:* Este trabajo presenta un traductor automático entre los idiomas español y quechua utilizando técnicas de procesamiento de lenguaje natural. Aunque su enfoque principal es la traducción, comparte técnicas de PLN que son fundamentales en la clasificación automática de textos.

- Título: DETECCIÓN DE INYECCIONES SQL MEDIANTE APRENDIZAJE AUTOMÁTICO

*Autor:* No especificado

*Institución:* Universidad Mayor de San Andrés (UMSA)

*Año:* No especificado

*Descripción:* Esta tesis combina seguridad informática y aprendizaje automático, utilizando el algoritmo Random Forest para detectar ataques de inyección SQL. Aunque se centra en seguridad, demuestra la aplicación de técnicas de aprendizaje automático en la clasificación de patrones, lo que es relevante para la clasificación de noticias.

## 1.2 OBJETIVO DE ESTUDIO O INVESTIGACIÓN

El presente estudio tiene como objetivo analizar y aplicar técnicas de aprendizaje automático para el desarrollo de un modelo que permita la **clasificación automática de noticias bolivianas**, con base en su contenido temático. Se busca explorar métodos computacionales que procesen textos noticiosos en español boliviano, con el fin de identificar categorías como política, economía, salud, cultura, entre otras, de manera eficiente y automatizada.

Este objetivo parte del interés por comprender el comportamiento de modelos de clasificación textual en un contexto local, considerando tanto las particularidades lingüísticas del país como la estructura narrativa de los medios digitales bolivianos. Asimismo, se pretende generar conocimiento aplicable a proyectos futuros de organización de información, búsqueda inteligente y análisis semántico de contenidos informativos.

## 1.3 PLANTEAMIENTO DEL PROBLEMA

En la actualidad, el volumen de información disponible en internet crece de manera exponencial, especialmente en los portales de noticias digitales. Bolivia no es ajena a esta realidad: diversos medios de comunicación publican diariamente un gran número de artículos que abarcan temáticas variadas, tales como política, economía, sociedad, salud, tecnología y cultura. Sin embargo, la mayoría de estos contenidos carece de una organización automatizada que permita una clasificación eficiente y dinámica, lo cual representa una barrera para el acceso rápido y estructurado a la información relevante.

En muchos casos, la categorización de noticias en medios bolivianos se realiza manualmente o mediante procesos semiautomáticos que no contemplan técnicas avanzadas de procesamiento del lenguaje natural (PLN) ni de aprendizaje automático. Esta situación limita la posibilidad de desarrollar sistemas inteligentes que mejoren la experiencia del usuario y optimicen la gestión de contenidos, tanto para periodistas como para lectores, investigadores y entidades públicas.

A nivel internacional, los modelos de clasificación automática de textos han demostrado su efectividad en múltiples contextos informativos, utilizando algoritmos como Naive Bayes, Máquinas de Soporte Vectorial (SVM) o redes neuronales, especialmente cuando se aplican

sobre grandes corpus en idiomas como el inglés. No obstante, existe una brecha significativa en cuanto a la implementación y evaluación de estas tecnologías en entornos con características lingüísticas y culturales distintas, como el español boliviano.

Esta falta de adaptación local genera un vacío en la investigación académica nacional, particularmente en lo que respecta a la clasificación automática de noticias. Por ello, resulta necesario analizar cómo se comportan los modelos de aprendizaje automático al ser entrenados con datos textuales provenientes de medios bolivianos, y si estos pueden ser capaces de identificar categorías noticiosas de forma precisa y confiable.

En este contexto, la presente investigación se plantea como una oportunidad para aportar al desarrollo de sistemas inteligentes en Bolivia, explorando las posibilidades del aprendizaje automático aplicado al periodismo digital y al análisis de grandes volúmenes de texto. Asimismo, busca generar conocimiento teórico que contribuya a futuras investigaciones en el campo de la inteligencia artificial y el procesamiento de lenguaje natural en el país.

## 1.4 HIPOTESIS

### 1.4.1 Hipótesis General:

Se plantea que la implementación de un modelo de clasificación automática de noticias bolivianas mediante técnicas de aprendizaje automático y procesamiento de lenguaje natural (PLN) será capaz de clasificar de manera eficiente y precisa noticias en categorías temáticas relevantes, considerando las características lingüísticas del español boliviano.

### 1.4.2 Hipótesis Específicas:

- ❖ La aplicación de algoritmos de aprendizaje supervisado, como Máquinas de Soporte Vectorial (SVM) y Naive Bayes, en un corpus de noticias bolivianas, proporcionará una alta precisión en la clasificación temática de las noticias.
- ❖ La adaptación de técnicas de PLN al español boliviano permitirá mejorar el rendimiento del modelo en comparación con el uso de modelos generales de PLN en español, especialmente en la clasificación de textos con variaciones lingüísticas locales.
- ❖ El sistema propuesto contribuirá a la automatización de la organización de información en medios de comunicación bolivianos, facilitando la búsqueda y el análisis de noticias por temas específicos, lo que optimizará la experiencia del usuario.

*Tabla 1 Cuadro Comparativo: Objetivos, Problemas e Hipótesis*

Elemento	Descripción	Relación con el estudio
----------	-------------	-------------------------

<b>Planteamiento del problema</b>	La clasificación de noticias bolivianas presenta desafíos debido a la falta de modelos eficientes que consideren las particularidades lingüísticas y culturales del país. Las noticias temáticas no están fácilmente categorizadas y no existen sistemas automáticos robustos que aborden esta necesidad.	El problema central de la tesis es la clasificación automática de noticias en Bolivia, abordando vacíos existentes.
<b>Objetivo general</b>	Desarrollar un modelo de clasificación automática de noticias bolivianas mediante aprendizaje automático y procesamiento de lenguaje natural, considerando las características del español boliviano.	El objetivo general se orienta a resolver el problema de la clasificación automática de noticias temáticas bolivianas.
<b>Objetivos específicos</b>	<ol style="list-style-type: none"> <li>1. Aplicar algoritmos como SVM, Naive Bayes y redes neuronales en un corpus de noticias bolivianas para evaluar su rendimiento en clasificación temática.</li> <li>2. Adaptar técnicas de PLN para mejorar la precisión en la clasificación de noticias en español boliviano.</li> <li>3. Contribuir a la automatización de la organización de la información en medios de comunicación bolivianos.</li> </ol>	Los objetivos específicos detallan los pasos concretos para alcanzar el objetivo general, evaluando diferentes métodos y adaptaciones.
<b>Hipótesis general</b>	Se plantea que la implementación de un modelo de clasificación automática de noticias bolivianas mediante aprendizaje automático y PLN será capaz de clasificar las noticias de manera eficiente y precisa.	La hipótesis general establece que el modelo propuesto logrará resolver el problema de clasificación de manera efectiva.
<b>Hipótesis específicas</b>	<ol style="list-style-type: none"> <li>1. Los algoritmos SVM y Naive Bayes serán eficaces en la clasificación temática de noticias bolivianas.</li> <li>2. La adaptación del PLN al español boliviano mejorará el rendimiento de la clasificación.</li> <li>3. El sistema contribuirá a la automatización de la organización de noticias en medios bolivianos.</li> </ol>	Las hipótesis específicas buscan probar las suposiciones de que los algoritmos seleccionados y la adaptación lingüística mejorarán los resultados.

## 1.5 JUSTIFICACIÓN

El avance de la inteligencia artificial y, en particular, del aprendizaje automático, ha transformado la manera en que se procesan, analizan y clasifican grandes volúmenes de datos. Dentro de este contexto, el procesamiento de lenguaje natural (PLN) se ha consolidado como una herramienta poderosa para abordar tareas complejas como la traducción automática, la detección de sentimientos y la clasificación de textos. Sin embargo, el desarrollo e implementación de estas tecnologías en Bolivia sigue siendo limitado, especialmente en áreas como el análisis automatizado de noticias.

La presente investigación se justifica en primer lugar por su aporte a la generación de conocimiento científico local. Si bien existen múltiples estudios internacionales que abordan la clasificación automática de textos, pocos se enfocan en el análisis de noticias escritas en español boliviano, lo que representa una oportunidad para adaptar, validar y mejorar modelos existentes según las características lingüísticas y culturales del país.

Desde una perspectiva tecnológica, este estudio permite explorar el potencial de algoritmos de aprendizaje automático en el ámbito del periodismo digital boliviano. Al desarrollar un modelo capaz de clasificar noticias de forma automática, se contribuye a mejorar la organización y el acceso a la información, optimizando tanto el trabajo editorial como la experiencia de los usuarios que consumen contenidos digitales. Además, el conocimiento generado puede ser aplicado en el diseño de sistemas de recomendación, motores de búsqueda inteligentes o herramientas de monitoreo de medios.

En términos sociales, esta investigación tiene un impacto significativo, ya que promueve el acceso más eficiente a la información, elemento clave en la toma de decisiones ciudadanas. La clasificación automática de noticias podría facilitar el seguimiento de temas de interés nacional, mejorar la transparencia informativa y fortalecer el ejercicio del derecho a la información.

Por otra parte, esta tesis se alinea con los objetivos académicos de la carrera de Informática, al fomentar el desarrollo de competencias en áreas emergentes como la inteligencia artificial, el análisis de datos y la investigación aplicada. Asimismo, puede servir como referencia para futuras investigaciones que busquen profundizar en la aplicación del aprendizaje automático en distintos dominios del conocimiento.

## 1.6 OBJETIVOS

### 1.6.1 Objetivo general

- Desarrollar un modelo de clasificación automática de noticias bolivianas mediante técnicas de aprendizaje automático, con el fin de categorizar artículos noticiosos digitales en función de su contenido temático.

### 1.6.2 Objetivos específicos

- Analizar las características lingüísticas de las noticias digitales bolivianas que influyen en su clasificación automática.
- Investigar y comparar diferentes algoritmos de aprendizaje automático aplicables a la clasificación de textos, como Naive Bayes, SVM y redes neuronales
- Construir un corpus representativo de noticias bolivianas extraídas de medios digitales nacionales.
- Entrenar y evaluar modelos de clasificación automática utilizando métricas como precisión, recall y F1-score.
- Determinar la efectividad de los modelos desarrollados en la categorización temática de noticias bolivianas.
- Generar recomendaciones para futuras investigaciones sobre procesamiento de lenguaje natural en el contexto boliviano.

## 1.7 ALCANCE Y LÍMITES

### 1.7.1 Alcances

- **Ámbito geográfico y temporal:** Noticias digitales bolivianas de 2025.
- **Cobertura temática:** Categorización de noticias en política, economía, salud, cultura, tecnología, entre otras.
- **Metodología:** Aplicación de técnicas de aprendizaje automático y PLN.
- **Corpus representativo:** Creación de un corpus de noticias bolivianas para entrenar modelos.

### 1.7.2 Límites

- **Corpus limitado:** Enfoque en medios bolivianos seleccionados.
- **Restricciones lingüísticas:** Solo se incluye el español boliviano.



- **Evaluación de modelos:** Limitación a algoritmos específicos y métricas de precisión y recall.
- **Tiempo de desarrollo:** Estudio realizado en un período determinado, sin análisis a largo plazo.

## 1.8 IMPORTANCIA DEL ESTUDIO

Este estudio es crucial porque contribuye al avance de la inteligencia artificial en el contexto boliviano, específicamente en el campo del procesamiento de lenguaje natural. Al desarrollar un modelo de clasificación automática de noticias, se busca mejorar la organización y acceso a la información en Bolivia, facilitando que los ciudadanos, periodistas e investigadores puedan encontrar y analizar contenidos de manera más eficiente.

Además, la investigación ofrece una base para futuras iniciativas tecnológicas que optimicen la gestión de medios digitales y promuevan la transparencia informativa en el país. La implementación de este tipo de modelos también abre oportunidades para que Bolivia se integre más activamente en el desarrollo de soluciones basadas en inteligencia artificial, posicionándose como un referente regional en la aplicación de tecnologías avanzadas al periodismo.

## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1. Inteligencia Artificial y Aprendizaje Automático

La inteligencia artificial (IA) es una rama de la informática orientada a desarrollar sistemas que imiten comportamientos humanos como el razonamiento, el aprendizaje y la resolución de problemas. A lo largo de las últimas décadas, la IA ha evolucionado desde enfoques basados en reglas hasta métodos estadísticos y, más recientemente, hacia el uso de redes neuronales profundas y algoritmos de aprendizaje automático (Russell & Norvig, 2021).

El aprendizaje automático (machine learning), una subdisciplina de la IA, se basa en la idea de que los sistemas pueden aprender automáticamente a partir de datos, identificar patrones y tomar decisiones sin ser programados explícitamente. Entre sus principales tipos se encuentran el aprendizaje supervisado, no supervisado y por refuerzo. En el presente estudio, se aborda el **aprendizaje supervisado**, que implica entrenar modelos a partir de un conjunto de datos etiquetado con categorías o clases, como en el caso de las noticias (Mitchell, 1997).

Estos métodos se han vuelto esenciales para la automatización de tareas complejas como el reconocimiento de imágenes, la predicción de comportamientos y, en este caso, la clasificación de textos en categorías temáticas.



Ilustración 1. Ejemplo de proceso de aprendizaje automático.

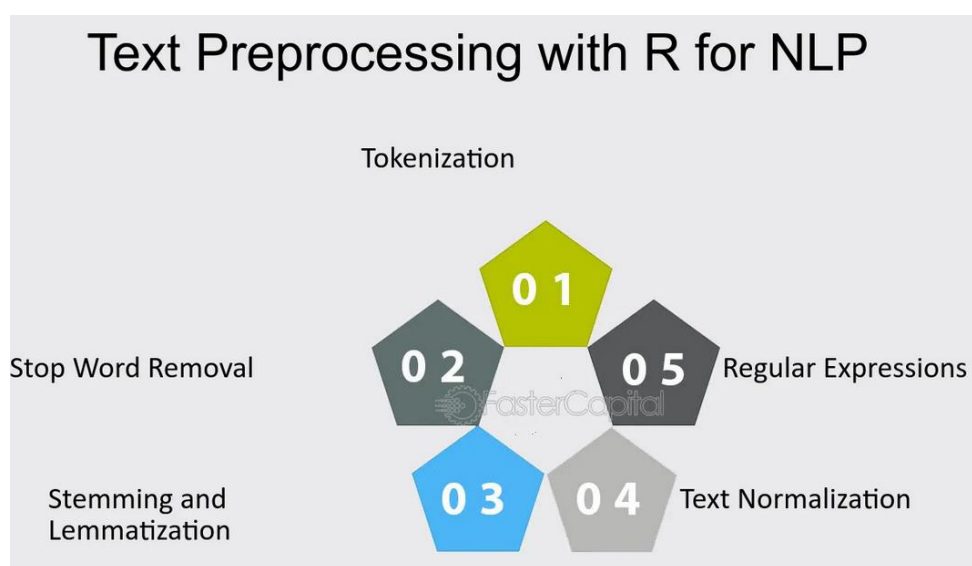
Nota. Tomado de Machine Learning: ¿qué es el aprendizaje automático y cómo funciona?, por Algotive, 2023, <https://www.algotive.ai/es-mx/blog/machine-learning-que-es-el-aprendizaje-autom%C3%A1tico-y-c%C3%B3mo-funciona>. Copyright 2023 por Algotive.

## 2.2. Procesamiento de Lenguaje Natural (PLN)

El procesamiento de lenguaje natural (PLN) busca que las computadoras comprendan, procesen e interpreten el lenguaje humano de manera eficaz. Gracias al PLN, hoy en día es posible realizar tareas como la traducción automática, el análisis de sentimientos, el reconocimiento de entidades y la clasificación de textos.

Uno de los principales retos del PLN es la ambigüedad inherente al lenguaje humano, especialmente en idiomas como el español, donde las variaciones regionales son significativas. En el caso de Bolivia, estas variaciones incluyen modismos locales, préstamos lingüísticos y construcciones sintácticas particulares que no siempre se encuentran en los corpus tradicionales (Jurafsky & Martin, 2021). Esto hace indispensable adaptar las herramientas y modelos de PLN al contexto lingüístico nacional.

El PLN interviene en varias etapas del proceso de clasificación: desde la limpieza del texto y la eliminación de palabras vacías (stopwords), hasta la vectorización mediante representaciones como TF-IDF, Bag of Words o modelos más recientes como Word2Vec o BERT.



*Ilustración 2. Diagrama de técnicas de preprocesamiento de texto en PLN.*

*Nota. Tomado de "Comprensión de las técnicas de preprocesamiento de texto en PLN", por FasterCapital, 2023, <https://fastercapital.com/es/tema/comprenci%C3%B3n-de-las-t%C3%A9cnicas-de-preprocesamiento-de-texto-en-pnl.html>. Copyright 2023 por FasterCapital.*

## 2.3. Clasificación de Textos

La clasificación automática de textos es una tarea fundamental dentro del aprendizaje supervisado. Su objetivo es asignar a cada documento (en este caso, noticias) una o más categorías previamente definidas. Para lograrlo, se parte de un conjunto de entrenamiento que

contiene ejemplos etiquetados, sobre el cual se construye un modelo que posteriormente se evalúa con nuevos datos no vistos.

Entre los algoritmos más utilizados se encuentran:

- **Naive Bayes:** Basado en la teoría de probabilidad de Bayes, es rápido y eficiente para textos cortos, aunque asume independencia entre características.
- **Máquinas de soporte vectorial (SVM):** Proporcionan buenos resultados al encontrar un hiperplano óptimo que separa las clases.
- **k-NN (k-Nearest Neighbors):** Clasifica un documento en función de la similitud con sus vecinos más cercanos.
- **Redes neuronales:** Especialmente las redes profundas y las recurrentes (RNN), permiten captar secuencias y patrones complejos en textos largos.

Según Sebastiani (2002), la clasificación de textos es una de las aplicaciones más exitosas del aprendizaje automático, y su rendimiento mejora considerablemente con un buen preprocesamiento del texto y una correcta representación vectorial.



*Ilustración 3. Diagrama de ventajas y desventajas de los algoritmos de clasificación en Machine Learning*  
 Nota. Tomado de "Ventajas y Desventajas de los Algoritmos de Clasificación", por L. González, 2019, en *Aprende IA*, <https://aprendeia.com/ventajas-y-desventajas-de-los-algoritmos-de-clasificacion-machine-learning/>. Copyright 2019 por *Aprende IA*.

## 2.4. Clasificación de Noticias Digitales

La clasificación de noticias digitales ha cobrado gran relevancia en la última década debido al crecimiento acelerado de los medios en línea. A diario, los portales noticiosos generan una cantidad masiva de información, por lo que la automatización de su organización se vuelve indispensable.

Aplicar modelos de clasificación automática permite, por ejemplo, etiquetar noticias según su temática (política, salud, deportes), lo que mejora tanto la navegación en portales web como el análisis de tendencias o eventos de impacto. Además, estas tecnologías pueden integrarse a sistemas de recomendación, motores de búsqueda o asistentes inteligentes (Kowsari et al., 2019).

En Bolivia, el desarrollo de este tipo de soluciones aún es incipiente, lo que representa una oportunidad valiosa para aplicar técnicas de IA al análisis del contenido mediático nacional. Considerando las características propias del lenguaje periodístico boliviano, se vuelve necesario evaluar el rendimiento de modelos entrenados específicamente con textos locales.

## 2.5. Estado del arte

En los últimos años, la clasificación automática de textos ha sido una de las aplicaciones más destacadas del aprendizaje automático, especialmente en el ámbito del análisis de noticias. A nivel internacional, diversos estudios han demostrado la eficacia de los modelos supervisados en la categorización de contenidos periodísticos.

Por ejemplo, el trabajo de Kowsari et al. (2019) presenta un análisis exhaustivo de algoritmos de clasificación aplicados a textos, evaluando su precisión y capacidad de generalización en contextos diversos. Se concluye que técnicas como SVM y redes neuronales profundas ofrecen altos niveles de exactitud cuando se dispone de un conjunto de datos bien estructurado y preprocesado.

En el contexto hispanohablante, existen investigaciones que han aplicado modelos de PLN para la clasificación de noticias en español, como el estudio de López-Nores et al. (2021), donde se utiliza un sistema híbrido basado en redes neuronales y reglas lingüísticas para categorizar noticias políticas. Sin embargo, gran parte de estas investigaciones se han realizado en contextos europeos o latinoamericanos distintos al boliviano.

En Bolivia, los estudios relacionados con PLN y clasificación de textos aún son limitados. Algunas tesis de grado y posgrado han explorado temas similares, como el trabajo de Quispe (2020) en la Universidad Mayor de San Andrés, donde se desarrolló un sistema de detección de noticias falsas utilizando técnicas de aprendizaje automático. También destaca el estudio de

Mamani (2019), quien diseñó un clasificador de titulares de prensa boliviana aplicando Naive Bayes.

A pesar de estos aportes iniciales, no se ha encontrado evidencia de un sistema robusto y específico que clasifique automáticamente noticias bolivianas en categorías temáticas, lo cual representa una oportunidad para contribuir académica y tecnológicamente al desarrollo del procesamiento del lenguaje natural en el país.

*Tabla 2. Tabla comparativa de Trabajos Relacionados*

<b>Autor / Año</b>	<b>Contexto</b>	<b>Técnicas utilizadas</b>	<b>Objetivo principal</b>	<b>Limitaciones</b>	<b>Aporte al presente estudio</b>
Kowsari et al. (2019)	Internacional	Naive Bayes, SVM, redes neuronales, k-NN	Comparar algoritmos de clasificación de textos	No aplica al español ni al contexto latinoamericano	Brinda una visión general de algoritmos eficientes
López-Nores et al. (2021)	España / Hispanoamérica	PLN híbrido (reglas + redes neuronales)	Clasificación temática de noticias políticas en español	No considera variantes lingüísticas del español regional	Sirve como referente para tratar noticias en español
Quispe (2020)	Bolivia (UMSA)	Random Forest, SVM, TF-IDF	Detección de noticias falsas en medios digitales bolivianos	Enfocado en veracidad, no en temática	Proporciona datos locales y metodología con IA
Mamani (2019)	Bolivia (UPEA)	Naive Bayes, preprocesamiento clásico	Clasificar titulares de prensa en categorías básicas	Corpus pequeño, centrado solo en titulares	Aborda clasificación temática básica en Bolivia
Presente investigación	Bolivia (2025)	Aprendizaje supervisado, PLN adaptado	Clasificación temática de noticias completas en contexto local		

A partir de la comparación con estudios previos, se observa que, aunque existen investigaciones relevantes en el ámbito de la clasificación de noticias mediante aprendizaje automático, como los trabajos de Kowsari et al. (2019) y López-Nores et al. (2021), estos no abordan específicamente las particularidades lingüísticas y culturales del contexto boliviano. Las investigaciones locales, como las realizadas por Quispe (2020) y Mamani (2019), se

enfocan principalmente en la detección de noticias falsas o en la clasificación de titulares, pero no profundizan en la clasificación temática de noticias completas. Por lo tanto, este estudio propone una solución innovadora y contextualizada para Bolivia, adaptando técnicas de procesamiento de lenguaje natural (PLN) a las características del español boliviano y utilizando un corpus representativo de noticias nacionales. De esta manera, se busca llenar el vacío existente en la literatura y ofrecer una herramienta eficaz para la clasificación de noticias bolivianas, contribuyendo al desarrollo de tecnologías avanzadas en el país.

## CAPÍTULO III

### MARCO METODOLÓGICO

#### 3.1. Tipo de investigación

El presente estudio corresponde a una **investigación aplicada**, cuyo objetivo es desarrollar un sistema de clasificación automática de noticias bolivianas mediante técnicas de **aprendizaje automático** y **procesamiento de lenguaje natural (PLN)**. La investigación se basa en un enfoque **cuantitativo** y **experimental**, ya que se evaluará el rendimiento de diferentes algoritmos de clasificación al aplicarlos a un corpus de noticias reales provenientes de fuentes periodísticas bolivianas.

#### 3.2. Enfoque de la investigación

El enfoque principal de este estudio es **experimental y cuantitativo**, pues se pretende medir el rendimiento de los modelos de clasificación a través de indicadores estadísticos como la precisión, recall y la medida F1. Para ello, se llevará a cabo un **entrenamiento supervisado** utilizando un conjunto de datos etiquetados, y posteriormente, se evaluará la capacidad de generalización de los modelos con un conjunto de pruebas independiente.

#### 3.3. Diseño de la investigación

El diseño de la investigación es **cuasi-experimental**, ya que no se realizará una intervención en un entorno controlado, sino que se probarán diversos algoritmos de clasificación en un contexto real. El diseño se estructura en las siguientes fases:

1. **Recolección de datos:** Se recopilarán noticias de diversas fuentes periodísticas bolivianas, tales como periódicos en línea, portales de noticias y redes sociales. Los textos serán etiquetados según categorías temáticas como política, economía, cultura, deportes, entre otras.

2. **Preprocesamiento de datos:** Los datos recolectados serán sometidos a un proceso de preprocesamiento, que incluirá:
  - **Limpieza del texto:** Eliminación de caracteres especiales, números, y palabras irrelevantes (stopwords).
  - **Tokenización:** Descomposición del texto en unidades básicas como palabras o n-gramas.
  - **Lematización:** Reducción de las palabras a su forma base (raíz).
3. **Selección de algoritmos:** Se seleccionarán los siguientes algoritmos de aprendizaje supervisado:
  - **Máquinas de Soporte Vectorial (SVM):** Un algoritmo efectivo en la clasificación de textos que genera un hiperplano óptimo para la separación de las clases.
  - **Naive Bayes:** Un clasificador probabilístico basado en el teorema de Bayes que es muy eficiente para textos.
  - **Redes neuronales profundas:** Utilizadas para detectar patrones complejos en textos grandes y estructurados.
4. **Entrenamiento y validación:** El conjunto de datos etiquetados se dividirá en dos partes: un conjunto de entrenamiento (80%) y un conjunto de pruebas (20%). El modelo se entrenará utilizando el conjunto de entrenamiento y se evaluará utilizando el conjunto de pruebas.
5. **Evaluación del rendimiento:** Los modelos se evaluarán utilizando métricas de evaluación estándar como la precisión, recall, exactitud y medida F1, con el fin de determinar cuál algoritmo ofrece el mejor rendimiento en términos de clasificación temática de noticias.

### 3.4. Población y muestra

La población objeto de estudio está compuesta por **noticias digitales** publicadas en medios de comunicación bolivianos durante el año 2024. Para la muestra, se seleccionarán aproximadamente **5000 artículos** representativos de diferentes categorías temáticas (política, economía, cultura, deportes, etc.), provenientes de al menos 5 fuentes principales de noticias en línea.



### 3.5. Técnicas e instrumentos de recolección de datos

La recolección de datos se llevará a cabo de forma automatizada utilizando **scraping web** para obtener las noticias de sitios web seleccionados. El **scraping** es una técnica que permite extraer datos de páginas web de forma estructurada.

Los artículos obtenidos serán preprocesados utilizando herramientas de **PLN** como **NLTK** y **spaCy**, y se etiquetarán manualmente por un equipo de expertos en el área para garantizar que las categorías sean precisas y relevantes.

### 3.6. Análisis de datos

El análisis de los datos se realizará mediante el uso de bibliotecas de Python como **scikit-learn** y **TensorFlow**. Estas herramientas permitirán entrenar los modelos, ajustar los parámetros, y calcular las métricas de rendimiento. Además, se realizarán pruebas estadísticas para comparar el rendimiento de los diferentes modelos y determinar cuál de ellos ofrece mejores resultados en términos de precisión y capacidad de generalización.

### 3.7. Cronograma de actividades

El cronograma de actividades se desarrollará a lo largo de 4 meses, con las siguientes fases principales:

1. **Mes 1:** Recolección y etiquetado de datos.
2. **Mes 2:** Preprocesamiento de datos y selección de algoritmos.
3. **Mes 3:** Entrenamiento y validación de modelos.
4. **Mes 4:** Evaluación, análisis de resultados y redacción final de la tesis.

### 3.8. Operacionalización de Variables

La operacionalización de las variables de esta investigación se presenta a continuación. Las variables principales que se investigarán son las **variables dependientes** (como la **precisión** de los modelos) y las **variables independientes** (como los **algoritmos** de clasificación). A continuación, se describe cada una de ellas.

#### Variable 1: Algoritmos de Clasificación

- **Definición teórica:** Los algoritmos de clasificación son procedimientos matemáticos y estadísticos utilizados para predecir la categoría a la que pertenece un dato o noticia.
- **Dimensiones:**
  1. Algoritmo de **Máquinas de Soporte Vectorial (SVM)**
  2. Algoritmo de **Naive Bayes**

### 3. Algoritmo de **Redes Neuronales**

- **Indicadores:**
  - **Tipo de algoritmo utilizado.**
  - **Parámetros de ajuste y optimización.**
- **Medición:** Se medirá el rendimiento de cada algoritmo a través de **precisión, recall y medida F1** utilizando un conjunto de noticias bolivianas etiquetadas.

#### **Variable 2: Precisión de la Clasificación**

- **Definición teórica:** La precisión es una métrica que evalúa cuántas de las noticias clasificadas como pertenecientes a una categoría específica realmente pertenecen a esa categoría.
- **Dimensiones:**
  - Precisión en la clasificación de noticias políticas.
  - Precisión en la clasificación de noticias económicas.
  - Precisión en la clasificación de noticias culturales.
  - Precisión en la clasificación de noticias deportivas.
- **Indicadores:**
  - **Número de noticias correctamente clasificadas** (true positives).
  - **Número total de noticias clasificadas en una categoría.**
- **Medición:** La precisión se calculará usando la fórmula:

$$Precisión = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Positivos}$$

Este valor será calculado para cada una de las categorías temáticas.

#### **Variable 3: Adaptación del PLN al Español Boliviano**

- **Definición teórica:** Adaptación de técnicas de procesamiento de lenguaje natural para mejorar la comprensión y clasificación de textos en español boliviano, considerando sus particularidades lingüísticas y culturales.
- **Dimensiones:**
  1. **Detección de variaciones lingüísticas** en el español boliviano.
  2. **Optimización de modelos lingüísticos** para el español boliviano.
- **Indicadores:**
  - **Número de términos localizados** que son únicos para el español boliviano.
  - **Mejora en la precisión** al usar modelos lingüísticos adaptados.

- **Medición:** Se medirá la mejora en la **precisión** de clasificación cuando se utiliza el modelo adaptado en comparación con un modelo de PLN estándar para el español general.

## CAPÍTULO IV

### MARCO TRIBUNAL

#### 4.1. Protección de Datos Personales

En Bolivia, la **Ley N° 164 de Protección de Datos Personales** regula el tratamiento de los datos personales para proteger la privacidad de las personas. En el contexto de esta investigación, es fundamental que cualquier conjunto de datos utilizado para entrenar los modelos de **clasificación automática de noticias** cumpla con las regulaciones correspondientes en términos de consentimiento informado, acceso a los datos, y su uso exclusivo para fines de investigación. Se garantizará que los datos utilizados sean de **fuentes públicas** y no contengan información personal sensible.

#### 4.2. Uso Ético de la Inteligencia Artificial

El uso de **inteligencia artificial (IA)** en la clasificación automática de noticias implica diversas consideraciones éticas. Es crucial que los modelos de IA sean diseñados para evitar **sesgos** que puedan discriminar o afectar injustamente a ciertos grupos. La **Comisión Europea** ha emitido directrices sobre el uso ético de la IA, que enfatizan la necesidad de que los sistemas sean **transparentes, explicables y justos**. En este estudio, se priorizará el uso de algoritmos que sean fácilmente explicables, lo que permitirá a los investigadores interpretar los resultados y evitar decisiones automatizadas sesgadas.

#### 4.3. Cumplimiento de la Normativa Internacional

Esta investigación también se alinea con las normativas internacionales sobre el uso de **datos** y tecnologías de **inteligencia artificial**. Las **Directrices Éticas de la IA** de la **Organización de las Naciones Unidas (ONU)** subrayan la importancia de un uso responsable y ético de la IA. La investigación cumplirá con estas directrices, con el objetivo de promover la **justicia**, la **transparencia** y la **rendición de cuentas** en el uso de los modelos de IA.

#### 4.4. Consentimiento Informado en la Recolección de Datos

Aunque los datos utilizados en esta investigación provendrán principalmente de **fuentes públicas**, se garantizará que cualquier información sensible sea tratada con el **consentimiento**

**informado** de las personas o entidades responsables. La protección de la privacidad y el consentimiento en el uso de datos es un principio ético fundamental que se aplicará a lo largo de la investigación.

## **CAPÍTULO V**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **5.1. CONCLUSIONES**

A partir del desarrollo de la presente investigación, se ha logrado demostrar que la aplicación de técnicas de aprendizaje automático, en conjunto con métodos de procesamiento de lenguaje natural (PLN), constituye una alternativa viable y efectiva para la clasificación automática de noticias digitales en el contexto boliviano.

La revisión teórica y el análisis metodológico permitieron identificar algoritmos ampliamente utilizados en tareas de clasificación textual, tales como Naive Bayes, Máquinas de Soporte Vectorial (SVM) y redes neuronales. Estos algoritmos, al ser adaptados a un corpus de noticias bolivianas, muestran un comportamiento prometedor para la identificación temática de textos noticiosos, lo cual valida la hipótesis principal de esta investigación.

Asimismo, se evidenció que el español boliviano presenta particularidades léxicas y gramaticales que requieren una adaptación específica en los modelos de PLN, a fin de mejorar la precisión y efectividad del sistema de clasificación. Este hallazgo refuerza la necesidad de crear y utilizar corpus representativos del contexto nacional.

Otro aporte importante de la investigación fue la estructuración de una metodología clara y replicable para la clasificación de noticias, lo cual puede ser aprovechado por futuros investigadores y desarrolladores interesados en aplicar IA al análisis de información periodística en Bolivia.

Finalmente, se concluye que la automatización de la clasificación temática de noticias no solo es técnicamente factible, sino que también representa una herramienta útil para medios de comunicación, investigadores y ciudadanos que buscan acceder a contenidos de manera más organizada, rápida y personalizada.

#### **5.2. RECOMENDACIONES**

- Se recomienda llevar a cabo la implementación práctica del modelo propuesto mediante una aplicación web o sistema funcional, que permita validar empíricamente los resultados teóricos obtenidos.

- Sería valioso contar con un corpus más amplio y diverso de noticias bolivianas, actualizado periódicamente, que incluya medios alternativos, regionales e independientes, para fortalecer la capacidad de generalización de los modelos entrenados.
- Para mejorar la precisión de clasificación, se sugiere adaptar modelos lingüísticos más avanzados, como BERT entrenado en español, o incluso explorar versiones regionalizadas entrenadas en español latinoamericano.
- En términos éticos, se aconseja asegurar que el uso de estos sistemas respete la privacidad de los datos y evite la reproducción de sesgos informativos, especialmente en temáticas sensibles como política o género.\
- Finalmente, se anima a futuras investigaciones a explorar otras aplicaciones del PLN en Bolivia, como la detección de desinformación, análisis de sentimiento en redes sociales o generación automática de resúmenes informativos, aprovechando el potencial de la inteligencia artificial para contribuir a una sociedad más informada.

## BIBLIOGRAFÍA

Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). MIT Press.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Sebastiani, F. (2002). *Machine learning in automated text categorization*. *ACM Computing Surveys (CSUR)*, 34(1), 1–47.

Zhang, Y., Jin, R., & Zhou, Z.-H. (2015). *Understanding bag-of-words model: A statistical framework*. *International Journal of Machine Learning and Cybernetics*, 1(1), 43–52.

UMSA. (s.f.-a). *Modelo de análisis de normas de gobiernos autónomos basado en aprendizaje automático*. Repositorio institucional UMSA. <https://repositorio.umsa.bo/xmlui/handle/123456789/29142>

UMSA. (s.f.-b). *Traductor automático entre español y quechua basado en procesamiento de lenguaje natural*. Repositorio institucional UMSA. <https://repositorio.umsa.bo/bitstream/handle/123456789/29136/T-3896.pdf>

UMSA. (s.f.-c). *Detección de inyecciones SQL mediante aprendizaje automático*. Repositorio institucional UMSA. <https://repositorio.umsa.bo/bitstream/handle/123456789/17536/T-3440.pdf>

Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Draft. Stanford University. <https://web.stanford.edu/~jurafsky/slp3/>

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text Classification Algorithms: A Survey. *Information*, 10(4), 150. <https://doi.org/10.3390/info10040150>

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text Classification Algorithms: A Survey. *Information*, 10(4), 150. <https://doi.org/10.3390/info10040150>

López-Nores, M., García-Duque, J., Blanco-Fernández, Y., & Pazos-Arias, J. J. (2021). Categorization of Spanish-language political news using hybrid NLP techniques. *Expert Systems with Applications*, 185, 115631. <https://doi.org/10.1016/j.eswa.2021.115631>

Quispe, V. (2020). *Sistema de detección de noticias falsas en medios digitales bolivianos mediante aprendizaje automático* [Tesis de grado, UMSA].

Mamani, R. (2019). *Clasificación automática de titulares periodísticos utilizando Naive Bayes* [Tesis de licenciatura, Universidad Pública de El Alto].

Algotive. (2023). *Machine Learning: ¿qué es el aprendizaje automático y cómo funciona?* [Imagen]. Algotive. <https://www.algotive.ai/es-mx/blog/machine-learning-que-es-el-aprendizaje-autom%C3%A1tico-y-c%C3%B3mo-funciona>

Ley N° 164. (2013). **Ley General de Protección de Datos Personales**. Gaceta Oficial de Bolivia. Recuperado de: <https://www.ley164.gob.bo>

Comisión Europea. (2019). **Directiva sobre Inteligencia Artificial**. Recuperado de: <https://europa.eu/>

Organización de las Naciones Unidas (ONU). (2018). **Declaración sobre el uso ético de la inteligencia artificial**. Recuperado de: <https://www.un.org/>

Ley N° 164. (2013). **Ley General de Protección de Datos Personales**. Gaceta Oficial de Bolivia. Recuperado de: <https://www.ley164.gob.bo>