

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERÍA Y ARQUITECTURA



Una Institución Adventista

Modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales: Una revisión sistemática de la literatura.

Por:

Estrella Sarit Barrientos Mogollon
Sol Abigail Mamani Mamani

Asesor:

Fredy Abel Huanca Torres

Lima, Diciembre de 2019

DECLARACIÓN JURADA DE AUTORÍA DEL TRABAJO DE INVESTIGACIÓN

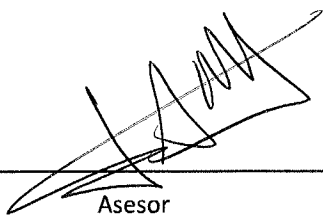
Mg. Fredy Abel Huanca Torres, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas , de la Universidad Peruana Unión.

DECLARO:

Que el presente trabajo de investigación titulado: *“Modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales: Una revisión sistemática de la literatura.”* constituye la memoria que presentan las estudiantes **Estrella Sarit Barrientos y Sol Abigail Mamani Mamani** para aspirar al Grado de Bachiller en Ingeniería de Sistemas, cuyo trabajo de investigación ha sido realizado en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente constancia en Lima, 04 diciembre del año 2019



Asesor

Mg. Fredy Abel Huanca Torres
DNI: 01345134

ACTA DE SUSTENTACIÓN DE TRABAJO DE INVESTIGACIÓN

En Lima, Naña, Villa Unión, a 2 día(s) del mes de diciembre del año 2019, siendo las 15:00 horas



se reunieron los miembros del jurado en la Universidad Peruana Unión Campus Lima, bajo la dirección del (de la) 362-4545
 presidente(a): Mg. Lizeth Geanina Huanca López
 secretario(a): Mg. Keyla Dervith De La Cruz Gutierrez y los demás miembros:
Ing. Jenson Daniel Chambi Aguilar, Mg. Nemias Saboya Ríos
 y el (la) asesor(a) MSc. Fredy Abel Huanca Torres

con el propósito de administrar el acto académico de sustentación del trabajo de investigación titulado: Modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales: Una revisión sistemática de la literatura

de los (las) egresados (as): a) Estrella Sarit Barrientos Mogollón
 b) Sol Abigail Mamani Mamani

conducente a la obtención del grado académico de Bachiller en Ingeniería de Sistemas
 (Denominación del Grado Académico de Bachiller)

El Presidente inició el acto académico de sustentación invitando a las candidato (a)/s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por las candidato (a)/s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato/a (a): Estrella Sarit Barrientos Mogollón

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	17	B+	Muy bueno	Sobresaliente

Candidato/a (b): Sol Abigail Mamani Mamani

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	17	B+	Muy bueno	Sobresaliente

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó a las candidato (a)/s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

[Firma]
 Presidenta/a
[Firma]
 Asesor/a
[Firma]
 Candidato/a (a)

[Firma]
 Miembro

[Firma]
 Secretario/a
[Firma]
 Miembro
[Firma]
 Candidato/a (b)

Modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales: Una revisión sistemática de la literatura.

Estrella Barrientos Mogollón¹, Sol Mamani Mamani²

¹ Ingeniería de sistemas, Universidad Peruana Unión, Perú

¹estrellabarrientos@upeu.edu.pe

²solmamani@upeu.edu.pe

Resumen. Las redes sociales se han convertido en la herramienta de comunicación e interacción más utilizada entre las personas y se han diversificado para cumplir funciones importantes dentro de la organización. En consecuencia, las redes sociales se han vuelto una fuente inmensa de datos que son procesados a través de modelos de aprendizaje supervisado para producir información que sea competente para la toma de decisiones como la predicción de campañas electorales, la predicción de consumo de un producto y/o servicio, la reputación de una empresa entre otros. De manera que el presente estudio tiene como objetivo identificar los modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales. Para la identificación de modelos de aprendizaje supervisado se realizó una revisión sistemática de la literatura (RSL) en bases de datos reconocidas y revistas indexadas. De un total de 1614 artículos se identificaron 32 artículos que hacen referencia a 6 modelos de aprendizaje supervisado y las funciones que cumplen como apoyo a la toma de decisiones en una organización. Se puede concluir que existen diversos modelos de aprendizaje supervisado siendo el de Support Vector Machine de mayor grado de precisión. También se han encontrado en las investigaciones modelos de: Naive Bayes, Decision Tree, Regression: Logistic y lineal, k-Nearest Neighbors, y finalmente Neural Network.

Palabras claves: Aprendizaje supervisado, Redes Sociales, Toma de decisiones

1 Introducción

En la actualidad el diario vivir se encuentra inmerso en la era de las tecnologías de información, cada día equipos avanzados y softwares inteligentes están a disposición de las personas permitiendo todo tipo de relacionamiento organizacional, empresarial o humano, gestionando la información y haciéndola accesible en todo momento. [1][5]

Las plataformas de redes sociales forman parte de estas tecnologías, y se han convertido en la herramienta de comunicación e interacción más utilizada entre las personas con el paso de los años [32]. De manera que son una fuente inmensa de conocimiento sobre las necesidades y comportamientos del cliente y son crucial para la gestión de las relaciones con los clientes [5]. Las empresas necesitan invertir en formar comunidades de redes sociales para supervisar constantemente el estado de su marca, compararlo con el de sus competidores y examinar periódicamente las medidas de mentalidad del cliente para guiar las decisiones de marketing [19].

La capacidad de las empresas y los gerentes para transformar estos datos en información valiosa podría marcar la diferencia entre el éxito y el fracaso del negocio [24]. El análisis de sentimiento es un subcampo del procesamiento de lenguaje natural (PLN) que permite identificar el contenido emocional de opiniones expresadas en un texto mediante el aprendizaje automático, de modo que se genere información precisa y actualizada de los usuarios de redes sociales con respecto a una empresa, producto, marca o evento [5][32]. Aquella información que tiene un valor comercial significativo y es una base para el proceso de toma de decisiones [5].

Obama fue uno de los primeros políticos en aplicar minería de opinión para analizar su partido político, clasificando las diferentes opiniones de las personas y prediciendo los resultados electorales de su campaña. Lo que llevó a muchas organizaciones a darse cuenta de la importancia de las redes sociales y de la manera que pueden usarse para producir importantes análisis y predicciones [20]

Los seres humanos continuamente tienden a querer predecir ¿Cómo votará la gente en las próximas elecciones?, ¿Ganará un equipo de fútbol si cambia su estrategia de juego?, ¿Subirá o bajará el mercado de valores?; la predicción es un proceso desafiante, pero que trae consigo recompensas potenciales y de mayor razón para las empresas. De manera que el presente trabajo de revisión sistemática tiene como objetivo identificar modelos de aprendizaje supervisado que sirvan de apoyo a la toma de decisiones organizacional basados en datos de las plataformas virtuales de redes sociales. Y por consiguiente, se busca contribuir a la investigación en áreas sociales y tecnológicas sobre el impacto que tiene la implementación de algoritmos inteligentes para procesar grandes volúmenes de datos, los cuales conservan los gustos, preferencias y opiniones de las personas.

Este artículo está distribuido de la siguiente manera: la sección II presenta el marco conceptual; la sección III describe la revisión sistemática de la literatura; la sección IV presenta los resultados de la revisión y finalmente la sección V describe las conclusiones y trabajos futuros.

2 Revisión de la literatura

En esta sección se presentan algunas definiciones del contexto sobre el cual se realiza el estudio y el objeto de análisis.

2.1 Inteligencia Artificial

La inteligencia artificial (IA) es una parte de la ciencia informática que se centra en la creación de máquinas y programas inteligentes. El propósito de la inteligencia artificial es tratar de imitar a la conciencia humana y realizar tareas como los seres humanos. En la práctica, significa la habilidad de una máquina o un programa para pensar y aprender, la inteligencia artificial se ha convertido en una parte importante de la industria de la tecnología. [25]

La IA ha tenido tres etapas a lo largo de su historia, entre 1950 y 1970, la llamada era de las redes neuronales, 1980-2000, la era del aprendizaje automático y la era del aprendizaje profundo en la actualidad. [25] Los factores más influyentes de la IA ahora son: la capacidad de computación y el poder de los datos y algoritmos. La potencia y capacidad de los procesadores han aumentado significativamente, hoy en día existe una enorme cantidad de datos disponibles que las máquinas pueden explotar, al mismo tiempo los costos de almacenamiento para la gestión de datos han disminuido y el desarrollo con almacenamiento de datos ha llevado a una forma más rápida de analizar cantidades masivas de datos. [26]

El papel de la IA en los procesos operativos de una organización ha cambiado significativamente en los últimos 30 años. Las empresas están experimentando un progreso en los niveles de madurez analítica que van desde el análisis descriptivo hasta el análisis predictivo y el aprendizaje automático. [27] Las empresas han tenido un éxito en el uso de IA para comprender dónde han estado y cómo puede aprender del pasado para participar en el futuro, pueden describir como diversas acciones y eventos afectarán los resultados, por lo cual requieren una forma de crear modelos predictivos que puedan reaccionar y cambiar cuando hay cambios en el entorno empresarial. [27]

2.1.1 Aprendizaje Automático

El aprendizaje automático es una ciencia que le permite a las máquinas desarrollar técnicas para poder aprender, este aprendizaje es un proceso automatizado que extrae patrones de los datos para construcción de modelos que permitan realizar la predicción utilizando algoritmos supervisados, relacionando características descriptivas actuales con características de destino basadas en un conjunto de instancias, consta de dos pasos: entrenamiento de la data. [28]

2.1.1.1 Aprendizaje supervisado

Infiere una función a partir de una serie de ejemplos etiquetados para posteriormente predecir una salida para otro conjunto distinto de ejemplos no etiquetados, de esta forma es como el algoritmo va aprendiendo a clasificar las entradas comparando con el modelo ya entrenado con sus etiquetas. [27]Algunos métodos y algoritmos del aprendizaje supervisado: Naive Bayes, Neural Network, Decision Tree, Logistic Regression, Support Vector Machine.[32]

2.1.1.1.1 Modelos de Aprendizaje Supervisado

Las personas han mediado sus experiencias en forma de contenidos que son almacenadas en redes sociales, lo que significa que generan grandes cantidades de datos que, teóricamente, pueden proporcionar información acerca de preferencias, opiniones, gustos, estados de ánimo, entre otros, de los usuarios. A causa del enorme tamaño de las redes y la gran cantidad de datos que son generados por ellas se requieren de técnicas y modelos especializados que generen una respuesta rápida con respecto a la petición hecha por el usuario final. [13]

La inteligencia artificial utiliza modelos inteligentes para predecir, clasificar y agrupar datos de manera que se obtenga información relevante para la toma de decisiones y para desarrollar flujos de trabajo rápidos, óptimos y prácticos para las personas. La función que realizan estos modelos es del entrenamiento, el cual necesita de datos de entrada para que el algoritmo utilizado en el modelo pueda aprender. Es decir, el término modelo de ML se refiere al artefacto de modelo que se crea en el proceso de entrenamiento.[33]

Los datos de entrenamiento deben contener la respuesta correcta, que se conoce como destino o atributo de destino. El algoritmo de aprendizaje encuentra patrones en los datos de entrenamiento que asignan los atributos de los datos de entrada al destino (la respuesta que desea predecir) y genera un modelo de aprendizaje automático que captura dichos patrones, de forma que los datos de entrenamiento ayuden a predecir las respuestas al tratar con datos nuevos. [33]

2.1.1.1.1.1 Naive Bayes

Al modelar sistemas inteligentes para aplicaciones del mundo real, inevitablemente hay que lidiar con la incertidumbre. Esta incertidumbre se debe a la imposibilidad de modelar todas las diferentes condiciones y excepciones que pueden ser la base de un conjunto finito de observaciones. La teoría de la probabilidad proporciona el marco matemáticamente consistente para cuantificar y calcular con incertidumbre. En principio, un modelo probabilístico asigna una probabilidad a cada uno de sus posibles estados. En los modelos para aplicaciones del mundo real, el número de estados es tan grande que una representación escasa del modelo es inevitable. Una clase general con una representación que permite modelar con muchas variables son las redes bayesianas.[27][28]

Las redes bayesianas están hoy bien establecidas como una herramienta de modelado para sistemas expertos en dominios con incertidumbre, las razones son su representación poderosa pero conceptualmente transparente para modelos probabilísticos en términos de una red.[28]

Por otro lado, dado que una red bayesiana define de manera única un modelo de probabilidad conjunta, la inferencia se basa en las reglas sólidas del cálculo de probabilidad.[28]

2.1.1.1.1.2 **Neural Network**

Son modelos matemáticos multivariantes no lineales que utilizan procedimientos iterativos, con el objetivo de minimizar una determinada función de error y clasificar así las observaciones. Las redes neuronales se componen de neuronas conectadas entre sí a través de nodos y capas. Dichas conexiones emulan las dendritas y los axones en los sistemas nerviosos biológicos, por los que se traspa información. Se emplean tanto en problemas supervisados como no supervisados, con la ventaja de que pueden separar regiones de forma no lineal. Su mayor desventaja es el carácter de “caja negra”, es decir, la dificultad en la interpretación de los resultados y la limitación a la hora de incorporar el sentido de negocio sobre la compleja estructura de pesos.[32]

2.1.1.1.1.3 **Support Vector Machine**

Son modelos de clasificación que tratan de resolver las dificultades que pueden suponer muestras de datos complejos, donde las relaciones no tienen por qué ser lineales. Es decir, se pretende clasificar a las observaciones en varios grupos o clases, pero estas no son separables vía un hiperplano en el espacio dimensional definido por los datos. Para ello, el conjunto de datos se embebe en un espacio de dimensión superior a través de una función 50 que permita poder separar los datos en el nuevo espacio a través de un hiperplano en dicho espacio. Entonces, se busca un hiperplano equidistante a los puntos más cercanos de cada clase (es decir, el objetivo es encontrar el hiperplano que separa las clases y que más dista de las observaciones de las mismas de forma simultánea).[31]

2.1.1.1.1.4 **Decision Tree**

Son técnicas de análisis que permite predecir la asignación de muestras a grupos predefinidos en función de una serie de variables predictivas. Los árboles de decisión son modelos sencillos y fácilmente interpretables, lo que hace que sean muy valorados por los analistas. Sin embargo, su poder predictivo puede ser más limitado que el de otros modelos, porque realizan una partición ortogonal del espacio, lo que convierte en silos la muestra y limita la capacidad predictiva debido a que este tipo de algoritmo tiende al sobreentrenamiento.[31]

2.1.1.1.1.5 **Linear Regression**

La regresión lineal es un modelo atractivo porque la representación es muy simple. La representación es una ecuación lineal que combina un conjunto específico de valores de entrada, cuya solución es la salida predicha para ese conjunto de valores de entrada diferente. Como tal, tanto los valores de entrada como el valor de salida son numéricos.

La ecuación lineal asigna un factor de escala a cada valor o columna de entrada, denominado coeficiente y representado por la letra griega mayúscula Beta (B). También se agrega un coeficiente adicional, que le da a la línea un grado adicional de libertad y a menudo se le llama intersección o coeficiente de polarización. En dimensiones superiores cuando tenemos más de una entrada (x), la línea se llama plano o hiperplano. Por lo tanto, la representación es la forma de la ecuación y los valores específicos utilizados para los coeficientes (por ejemplo, B0 y B1 en el ejemplo anterior). Es común hablar sobre la complejidad de un modelo de regresión como la regresión lineal. Esto se refiere al número de coeficientes utilizados en el modelo.[33]

2.2 Redes Sociales

Existe gran cantidad de definiciones acerca del concepto de red social “tradicional”, por lo que todas coinciden en que el núcleo fundamental para la creación y mantenimiento de las mismas son los miembros que las componen, ya sean personas físicas o no, y las relaciones que mantienen, por encima de sus atributos o características individuales; prima el tipo de relación y la expresión que en ella se desarrolle. Los cuales pueden ser de cualquier índole, manteniendo una serie de relaciones sociales, personales, físicas, religiosas, políticas y educativas. [21]

Debido al avance de la tecnología, telecomunicaciones y por supuesto la llegada del internet en el año 2000 a nivel mundial [19], se ha desarrollado herramientas que ayuden a las redes sociales a tener una mejor interacción entre los miembros que las componen, con la finalidad de establecer lazos virtuales con usuarios que incluso no conocen, no están geográficamente cerca suyo, no hablan su mismo idioma, sin embargo, tienen intereses en común. [21]

Con el surgimiento de las redes sociales como plataforma, día a día es mucha la información que se almacena proveniente de las mismas, organizaciones como Facebook, Twitter, LinkedIn y también privadas están obligadas a utilizar métodos que ayuden a manejar los volúmenes de datos que den una respuesta rápida, puntual y eficiente al usuario. [23] El gran crecimiento de los usuarios en las redes sociales las ha convertido en grandes fuentes de información para las empresas, las cuales invierten dinero para que se les otorgue data clasificada de acuerdo a su necesidad y giro de negocio. [23] En un principio las empresas no optaron por la idea de implementar los servicios que brindaban las redes sociales en sus actividades y ayuda de tomas de decisiones, ya que se trataba de un nuevo fenómeno que no dominaban, desconociendo el alcance que podía llegar a tener, pero gracias a la eficacia de la clasificación de data brindada por las redes sociales, sectores de industria, construcción, comercio y servicios externos empezaron a incluir en su plan de trabajo a la implementación de redes sociales como ayuda al marketing empresarial.[24]

La comprensión de las redes sociales evoluciona en un gran problema de datos cuando los especialistas en negocios, administración o sistemas de información esperan predecir el comportamiento para finalmente mejorar el marketing, las ventas y el comercio en línea. Uno de los problemas que afronta las redes sociales es clasificar la información que se busca, con el pasar del tiempo se desarrolló algoritmos que utilizan la información que proporciona la inteligencia colectiva para buscar o recomendar documentos. Debido a la necesidad del usuario el uso de la inteligencia artificial en las redes sociales ha ido creciendo y siendo un gran apoyo para este, notándose la mejor experiencia del usuario al navegar. La inteligencia artificial, es la facultad de comprender las relaciones entre los hechos y cosas. [20]

Uno de los enfoques de la inteligencia artificial son los agentes de software, son herramientas para la ayuda de la sobrecarga de información. Se define a un agente de software como un programa autocontenido capaz de controlar el proceso de toma de decisiones y de objetivos.

3 Método de la revisión sistemática de la literatura

El método de la revisión sistemática de la literatura tiene objetivo de resumir la información existente respecto a un tema de investigación en particular. Cuenta con una serie de etapas definidas en un protocolo de revisión específicos.

La realización de una revisión sistemática se justifica con el hecho de querer obtener información relevante y actualizada con un considerable grado de fiabilidad no teniendo que leer gran cantidad de artículos, debido a que la metodología brinda una serie de pasos que ayudan optimizar el proceso de investigación, haciéndola muchas más práctica. [14]

3.1 Necesidad de la revisión sistemática

La revisión sistemática de la literatura que se presenta en este estudio surge a partir de la necesidad de querer identificar qué modelos de Aprendizaje Supervisado existen como apoyo a la toma de decisiones en las organizaciones basado en los datos de redes sociales. Asimismo, se requiere identificar (2) ¿Qué funciones cumplen los modelos de aprendizaje supervisado como apoyo a la toma de decisiones organizacional basados en datos de redes sociales? y (3) ¿Qué herramientas se utilizaron en los modelos de aprendizaje supervisado encontrados?

TABLA 1. ELABORACIÓN DEL OBJETIVO DE LA INVESTIGACIÓN

Campos/criterios	Valor
-------------------------	--------------

Objeto de estudio	Aprendizaje Supervisado
Propósito	Identificar
Foco	Modelos, técnicas y métodos
Involucrados	Redes sociales, modelos de aprendizaje supervisado
Factores del contexto	Redes sociales

3.2 Preguntas para la revisión sistemática

Para la definición y estructuración de las preguntas de investigación se tomó como referencia la finalidad de la investigación expuesta en la sección anterior. En la siguiente Tabla 2 se muestran las preguntas de investigación y en la Tabla 3 las preguntas bibliométricas.

TABLA 2. PREGUNTAS DE INVESTIGACIÓN Y MOTIVACIÓN

IDs	Pregunta	Motivación
PI-01	¿Qué modelos de aprendizaje supervisado existen como apoyo a la toma de decisiones organizacional basados en datos de redes sociales?	Identificar los modelos de aprendizaje supervisado se utilizan como apoyo a la toma de decisiones en las organizaciones que son basados en datos de redes sociales.
PI-02	¿Qué funciones cumplen los modelos de aprendizaje supervisado como apoyo a la toma de decisiones organizacional basados en datos de redes sociales?	Identificar qué permiten realizar los modelos de aprendizaje supervisado que sirvan de apoyo a la toma de decisiones organizacional.
PI-03	¿Qué herramientas se utilizaron en los modelos de aprendizaje supervisado encontrados?	Identificar las herramientas utilizadas por los modelos de aprendizaje supervisado.

TABLA 3 . PREGUNTAS DE BIBLIOMETRIA

IDs	Pregunta	Motivación
PB-01	¿Cómo ha evolucionado en el tiempo la frecuencia de las publicaciones sobre este tema?	Identificar la evolución en el tiempo sobre las publicaciones sobre este tema
PB-02	¿Cuáles son las publicaciones en las que se han encontrado estudios relacionados al tema?	Identificar en qué dominio de aplicación se concentra la mayor cantidad de publicaciones sobre este tema.
PB-03	¿Cuál es la cantidad de publicaciones por tipo de artículo?	Determinar la cantidad de estudios publicados por tipo de artículo para identificar la concentración de los mismos.

3.3 Definición de las cadenas de búsqueda

Basado en la guía para la elaboración de la Revisión sistemática de la literatura (RSL) de Kitchenham [15], el protocolo de la revisión consiste en la especificación formal de los pasos a seguir durante la realización de la revisión sistemática. A continuación, se presentan los criterios utilizados para la ejecución de la búsqueda de la literatura relevante para la investigación a desarrollar en este estudio.

Cadena de Búsqueda:

A continuación se muestran las cadenas de búsqueda para la selección de resultado basado en la estrategia PICO. [16]

Población:

Entidad: Modelos de Aprendizaje Supervisado

Término principal 1: Modelos

Términos alternos: Métodos y técnicas.

Justificante: Se selecciona el objeto de estudio y los términos alternativos similares al término principal.

Término principal 2: Aprendizaje Supervisado

Términos alternos: Aprendizaje automático supervisado

Justificante: Se selecciona el término por ser el área de estudio a investigar y se identifica los términos alternos por ser sub área de estudio en este campo.

Intervención:

Entidad: Aplicado como apoyo a la toma de decisiones en las organizaciones basado en datos de redes sociales

Término principal 1: Aplicado como apoyo a la toma de decisiones en las organizaciones

Términos alternos: Empresas y Entidades

Justificante: Identificar el término principal en donde las modelos son utilizadas e identificar los términos similares al objeto principal.

Término principal 2: Redes sociales

Términos alternos: Social Media

Justificante: Se selecciona el término por ser el recurso principal de datos y se seleccionan los términos alternos por ser las más relaciones al término principal.

Comparación: No aplica debido a que no es parte del objetivo de investigación.

Resultado:

Entidad: Modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales

Término principal 1: Modelos de aprendizaje supervisado

Términos alternos: Métodos y técnicas de aprendizaje automático supervisado

Término principal 2: Apoyo a la toma de decisiones en las organizaciones

Términos alternos: Empresa y entidad

Término principal 3: Redes sociales

Términos alternos: Social media

Justificante: Se seleccionan dichos términos puesto que es lo que se busca obtener como resultado de la investigación.

Idioma:

El idioma elegido para definir la cadena de búsqueda ha sido el inglés. Luego siguiendo las recomendaciones de la estrategia PICO, se obtuvo como resultado la cadena de búsqueda a partir del uso de operadores booleanos entre los elementos definidos previamente: (Población) AND (Intervención) AND (Comparación) AND (Resultado). [16]

En la Tabla 4 se puede apreciar los elementos de la estrategia PICO a partir de los cuales se elabora la cadena de búsqueda.

TABLA 4. TÉRMINOS EN INGLÉS Y CONECTORES LÓGICOS A SER USADOS EN LA BÚSQUEDA

Concepto	Términos
Población	(Model* or Technique* or Method*) and (Supervised learning or Supervised machine learning)
Intervención	(Organization* or company* or entity*) and (Social network or Social Media)
Comparación	No aplica

Outcome	(Cases)
Contexto	Models of supervised learning

Tipos de Búsqueda:

Se desarrolló una búsqueda semi-automática en las librerías digitales previamente seleccionadas de acuerdo a su relevancia en el ámbito científico y al contexto que se requiere evaluar.

3.4 Criterios de inclusión y exclusión

De acuerdo a los lineamientos elaborados por Kitchenham[15], luego de ejecutar la cadena de búsqueda en las diferentes librerías indexadas, los resultados deben ser sometidos a evaluación para poder determinar cuáles son los estudios primarios que responden directamente las preguntas de investigación formuladas. En base a lo expuesto se tomó en consideración los siguientes criterios para la evaluación de los estudios:

TABLA 5. CRITERIOS DE INCLUSIÓN

Criterios de Inclusión	
C.I.1	Se consideran todos aquellos artículos provenientes de librerías digitales indexadas. (IEEE Xplore, ACM Digital Library , Science Direct, Springer Link)
C.I.2	Se consideran artículos cuya aplicación de modelo de aprendizaje supervisado haya tenido un factor de éxito.
C.I.3	Se consideran todos los artículos dentro del rango de temporalidad.
C.I.4	Se consideran solo los artículos que contengan estudios sobre los modelos de aprendizaje supervisado basados en datos de redes sociales.

TABLA 6. CRITERIOS DE EXCLUSIÓN

Criterios de Exclusión	
-------------------------------	--

C.E.1 Serán excluidos los artículos duplicados.

C.E.2 Serán rechazados los artículos de contenido similar, quedándose solo los que tengan el contenido más completo.

C.E.3 Serán excluidos los artículos cuyo título no tenga relación con el objeto de estudio.

C.E.4 Serán excluidos los estudios secundarios, estudios terciarios y resúmenes.

C.E.5 Serán excluidos los artículos que no se encuentren en inglés.

Temporalidad:

Debido a la evolución del uso de las redes sociales en los últimos 5 años, para el presente estudio se toma en consideración los estudios desarrollados desde el año 2015 hacia adelante. [22]

Fuentes de Datos:

Las librerías digitales indexadas consideradas por su relevancia científica para la selección de artículos fueron:

- ScienceDirect (<http://www.sciencedirect.com>)
- ACM Digital Library(<https://dl.acm.org/>)
- IEEE Xplore Digital Library(<https://ieeexplore.ieee.org/Xplore/home.jsp>)
- Springer Link (<https://link.springer.com/>)

Procedimientos para la selección de estudios:

Se considera el siguiente procedimiento para la selección de artículos en la revisión sistemática:

- **Etapa 1:** Se procedió a ejecutar la cadena de búsqueda PICO, en la base de datos indexadas previamente seleccionadas, aplicando los criterios de inclusión y exclusión de acuerdo a la Tabla VI. Las referencias de los artículos resultantes fueron guardadas para su posterior refinamiento.
- **Etapa 2:** Se revisaron los títulos de los artículos resultantes de la ejecución de la Etapa 1 excluyendo solamente los artículos que no se encuentren en idioma inglés.

- **Etapa 3:** Se revisaron los resúmenes de los artículos previamente seleccionados en la Etapa 2 para proceder con la exclusión de todos los artículos no tenga relevancia con nuestro objeto de estudio.
- **Etapa 4:** Se procedió con la realización de una revisión preliminar del contenido de los artículos seleccionados luego de la Etapa 3, para aplicar el criterio de inclusión vista en la Tabla 7.

TABLA 7. PROCEDIMIENTOS Y CRITERIOS DE INCLUSIÓN

Procedimiento	Criterio de selección
Etapa 1	C.I.1, C.I.3,C.E.4,C.E.1
Etapa 2	C.E.3,C.E.5.
Etapa 3	C.E.2,C.I.2
Etapa 4	C.I.4

3.5 Definición del Protocolo de Investigación

Esquema de evaluación de calidad de estudios:

Seguidamente se definió el esquema de evaluación de calidad según los lineamientos de Kitchenham[15] para evaluar la calidad de los estudio seleccionados.

Dentro del esquema se definió una lista de criterios con la finalidad de comprobar el cumplimiento de cada artículo. Cada criterio está acompañado de un puntaje basado en la escala de Rouhani[18], el cual consiste en los siguientes puntajes: Sí cumple (S) = 1, Cumple parcialmente (P) = 0.5 y No cumple (N) = 0. Los resultados serán presentados según el esquema de la Tabla 8:

TABLA 8. CRITERIOS DE EVALUACIÓN DE CALIDAD

N°	Criterio de evaluación
1	<p>¿El método seleccionado para llevar a cabo el estudio ha sido documentado apropiadamente?</p> <p>S: El método seleccionado ha sido documentado apropiadamente. P: El método seleccionado ha sido documentado parcialmente. N: No se ha documentado el método seleccionado</p>

-
- 2 **¿Se han documentado las limitaciones del estudio de manera clara?**
 S: Las limitaciones se han documentado claramente.
 P: Las limitaciones se han documentado parcialmente.
 N: No se han documentado limitaciones.
-
- 3 **¿Los aportes del estudio comunidades científicas académicas o para la industria ha sido descritos?**
 S: Los aportes del estudio han sido mencionados claramente.
 P: Los aportes del estudio han sido mencionados claramente.
 N: No se han mencionado aportes.
-
- 4 **¿Los resultados han contribuido a responder las preguntas de investigación planteadas?**
 S: Los resultados han contribuido a responder todas las preguntas de investigación.
 P: Los resultados han contribuido a responder algunas preguntas de investigación.
 N: Los resultados no han contribuido a responder las preguntas de investigación.
-

Estrategia para la extracción de datos:

Se elaboró un formulario de extracción de datos de manera que se tenga un instrumento establecido que sirva de guía al momento de extraer información de los artículos seleccionados, como se detalla en la Tabla 9 en base a las pautas dadas por Kitchenham y Brereton [15], ello para que facilitar la resolución de preguntas de investigación.

TABLA 9. FORMULARIO DE REFERENCIA PARA LA EXTRACCIÓN DE DATOS

Criterio	Detalle	Relevancia
Identificador		
Fuente		
Título		
Autores		

Publicación
Año de Publicación
Tipo de publicación
Modelos de AS
Funciones

Validar el protocolo de investigación:

El protocolo utilizado para el desarrollo de la RSL, fue revisado por el Magister Scientiae Fredy Abel Huanca Torres.

4 Resultados

4.1 Resultados de la búsqueda

A continuación, se ejecutan las cadenas de búsqueda de acuerdo a los pasos definidos en la sección III. En la tabla 10 se pueden visualizar los resultados. Cabe resaltar que debido a la cantidad de información se optó por filtrar por sintaxis de la librería.

TABLA 10. RESULTADOS DE BÚSQUEDA

Base datos	Fecha	Total
Cadena de Búsqueda		
ACM Digital Library	Noviembre 2019	55
<pre>"query": { ("model\$" OR "technique\$" OR "method\$") AND ("supervised learning" OR "supervised machine learning") AND ("organization\$" OR "company\$" OR "entity\$") AND ("social network" OR "social media") } "filter": {"publicationYear":{"gte":2015 }}, {owners.owner=HOSTED}</pre>		
Science Direct	Noviembre 2019	830

IEEE Xplore Digital Library	41	34	20	16	10
Springer Link	668	529	23	9	4
ACM Digital Library	55	27	12	9	8
Total:	1614	1255	86	48	32

4.3 Evaluar calidad de los estudios

Los criterios de calidad fueron aplicados como último filtro para seleccionar los artículos, fueron 10 investigaciones en total que lograron una calificación mayor al 70 % de la puntuación total, lo cual se puede considerar como un buen indicador de la calidad de los estudios seleccionados para la RSL.

TABLA 12. EVALUACIÓN DE CALIDAD DE ESTUDIOS

ID	C1	C2	C3	C4	Total
1	0,5	1	0,5	0,5	2,5
2	0,5	1	0,5	0,5	2,5
3	1	1	1	1	4
4	1	1	1	1	4
5	1	1	1	1	4
6	0,5	0,5	0,5	1	2,5
7	0,5	0,5	1	0,5	2,5
8	1	0,5	0,5	1	3
9	0,5	1	1	0,5	3
10	0,5	1	1	1	3,5
11	0,5	0,5	0,5	1	2,5

12	0,5	1	0,5	0,5	2,5
13	1	1	0,5	0,5	2
14	0,5	0,5	1	1	3
15	1	0,5	0,5	1	3
16	1	0,5	0	1	2,5
17	0,5	0,5	0,5	1	2,5
18	1	1	0,5	1	3,5
19	1	1	0,5	1	3,5
20	1	0,5	0,5	1	3
21	0,5	0,5	1	1	3
22	1	1	1	1	4
23	1	0,5	0,5	0,5	2,5
24	1	0,5	0,5	1	3
25	0,5	1	0,5	0,5	2
26	1	0,5	1	1	3,5
27	1	1	1	1	4
28	0,5	1	0,5	0,5	2,5
29	0,5	1	1	0,5	3
30	1	1	1	1	4
31	1	1	0,5	1	3,5
32	1	1	0,5	0,5	3

4.4 Extraer resultados relevantes

De acuerdo con lo descrito en la guía de Kitchenham [15], los formularios para la extracción de datos deben ser diseñados con la finalidad de recolectar toda la información necesaria para resolver las preguntas de investigación del estudio.

Cada uno de los artículos seleccionados fue leído, estudiado y detallado en un el formulario planteado anteriormente, el cual se realizó en el mismo idioma del artículo. El detalle de los criterios para los cuales no se encontró información relevante fue llenado con las siglas NI (No se encontró información).

TABLA 13. EJEMPLO DE EXTRACCIÓN DE DATOS DE UN ESTUDIO PRIMARIO

Criterio	Detalle	Relevancia
Identificador	-	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Sentiment Analysis of Facebook Posts: the Uber case	PB-2
Autores	Anna Baj-Rogowska	PB-2
Publicación	The 8th IEEE International Conference on Intelligent Computing and Information Systems	PB-3
Año de publicación	2017	PB-1

Tipo de publicación	Conferencia	PB-2
Modelos de AS	N-vecinos, Árboles de decisión , Redes neuronales, Máquinas de vectores de soporte, Redes bayesianas, Regresión lineal, Regresión no lineal.	PI-1
Funciones	Analizar las opiniones en facebook sobre Uber	PI-2

4.5 Análisis bibliométrico

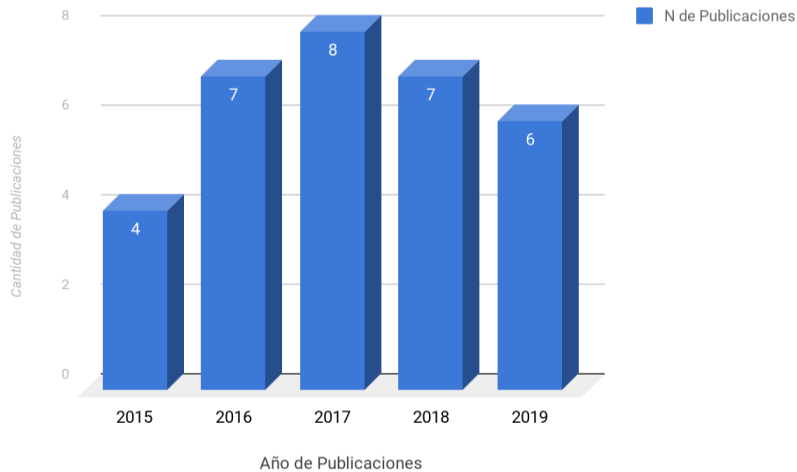
En esta sección se procede a describir el análisis de la tendencia de los artículos seleccionados para esta Revisión sistemática de la literatura(RSL) de acuerdo a factores como tiempo, tipo de artículo y tema tratado.

4.5.1 Pregunta de bibliometría 1(PB-1)

¿Cómo ha evolucionado en el tiempo la frecuencia de las publicaciones sobre este tema?

Como se puede ver en la Figura 1, un incremento en el número de publicaciones que describen los modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones en base a datos de redes sociales. De un total de de 32 artículos, 13(40.63%) han sido publicados en los últimos 2 años, 11(34.38%) durante el 2015 al 2016, y 8(25%) en el 2017. Con ello se demuestra que el año 2017 concentra la mayor frecuencia de artículos sobre modelos de aprendizaje supervisado en base a datos de redes sociales. Sin embargo se esperan muchas más publicaciones para este año 2019.

FIGURA 1. FRECUENCIA DE PUBLICACIONES



4.5.2 Pregunta de bibliometría 2(PB-2)

¿Cuáles son las publicaciones en las que se han encontrado estudios relacionados al tema?

En la Tabla 14 se presentan las publicaciones de donde se han extraído los artículos seleccionados. A partir de este análisis se puede observar que existen una gran variedad de dominios de aplicación sobre los modelos de aprendizaje supervisado, sin embargo se encontró una mayor frecuencia en las publicación de :International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Computers in Human Behavior y Tourism Management.

TABLA 14. PUBLICACIONES CORRESPONDIENTES A LOS ARTÍCULOS SELECCIONADOS

Publicación	Cantidad
International Conference on Advances in Social Networks Analysis and Mining (ASONAM)	02
23rd International Conference on High Performance Computing Workshops (HiPCW)	01
International Joint Conference on Neural Networks (IJCNN)	01

3rd International Conference on Computer Science and Engineering (UBMK)	01
The 8th IEEE International Conference on Intelligent Computing and Information Systems	01
European Conference on Electrical Engineering and Computer Science (EECS)	01
IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)	01
International Conference on Current Trends towards Converging Technologies (ICCTCT)	01
5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)	01
International Conference on Advanced Computer Science and Information Systems (ICACSIS)	01
9th International Conference on Social Media and Society	01
ACM Digital Library Transactions on the Web (TWEB)	01
25th ACM International on Conference on Information and Knowledge Management	01
22nd Pan-Hellenic Conference on Informatics	01
ACM Transactions on Knowledge Discovery from Data (TKDD)	01

21st International Database Engineering & Applications Symposium	01
Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory	01
International Journal of Information Management	01
7th International Conference on Advances in Computing & Communications	01
20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems	01
Computers in Human Behavior	02
Journal of Biomedical Informatics: X	01
Journal Industrial Marketing Management	01
Tourism Management	02
Requirements Engineering	01
Springer Link Journal of Big Data	01
Neural Computing and Applications	01

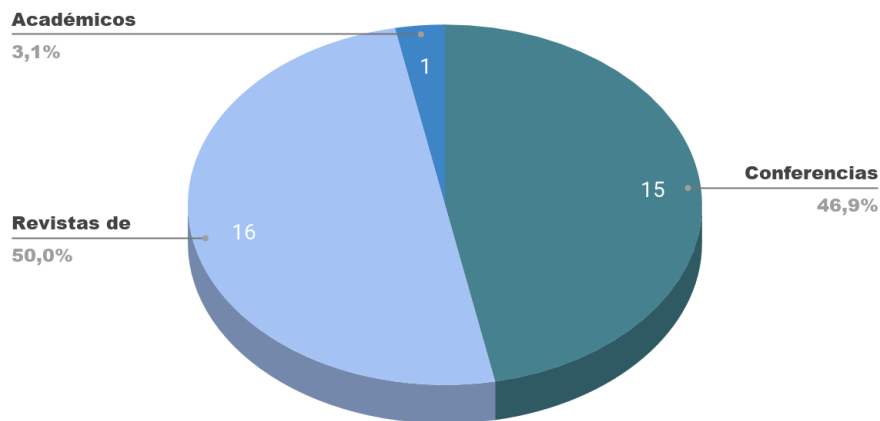
AI & SOCIETY	01
2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI	01

4.5.3 Pregunta de bibliometría 3(PB-3)

¿Cuál es la cantidad de publicaciones por tipo de artículo?

Como se observa en la Figura 2, los artículos de conferencia representan el 46.9% del total de artículos seleccionados para esta RSL, seguidamente se tiene a los artículos en revista con un 50% y los artículos académicos con un 3.13%. De este análisis se puede concluir que las conferencias y las revistas son la mayor fuente de datos de los estudios sobre modelos de aprendizaje supervisado.

FIGURA 2. CANTIDAD DE PUBLICACIONES POR TIPO



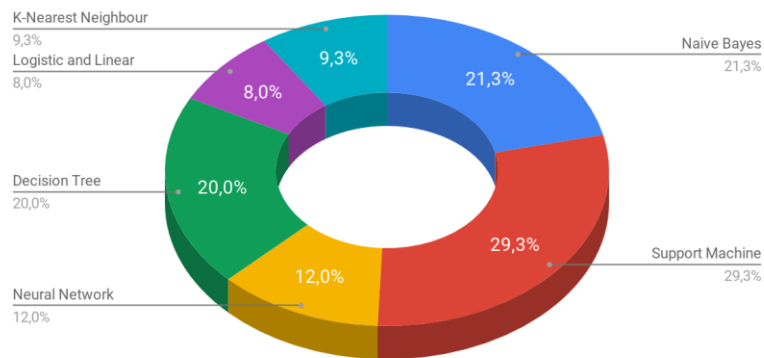
4.6 Sintetizar los datos extraídos

4.6.1 Pregunta de Investigación 1 (PI-1)

¿Qué modelos de aprendizaje supervisado existen como apoyo a la toma de decisiones en las organizaciones basados en la data de redes sociales?

A través de la información de los artículos, se pudo encontrar diferentes modelos de aprendizaje supervisado que ayudan a predecir futuros escenarios en las organizaciones, el modelo más utilizado es el de Máquinas de Vector Soporte con 28,6%, siendo caracterizada por la eficacia aun en espacios de grandes dimensiones, seguido por el modelo de Redes Bayesianas con un 20.8%, se puede observar los modelos utilizados en los artículos en la Figura 3.

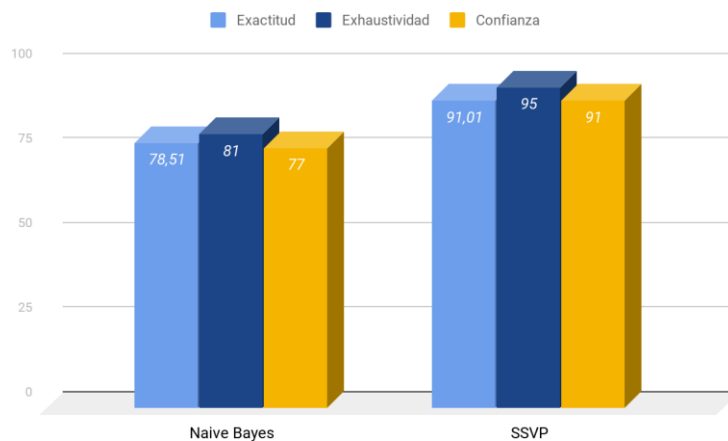
FIGURA 3. MODELOS SUPERVISADOS



La mayoría de los modelos de Aprendizaje Supervisado que se encontraron en esta investigación siguieron una serie de pasos para ser aplicados: la primera es la colección de datos, donde se tomaron datos de diferentes redes sociales a través de APIs, en este paso se examina la data para asegurar su calidad eliminando la data ruidosa y obteniendo una data limpia y pareja (implica en el porcentaje de precisión del modelo), en el siguiente paso se entrenó la data con diferentes algoritmos de diversos modelos, finalmente se evalúa la capacidad del modelo de predicción. A continuación se describirán algunos de los artículos más relevantes que encontramos de acuerdo al puntaje obtenido en los criterios de evaluación:

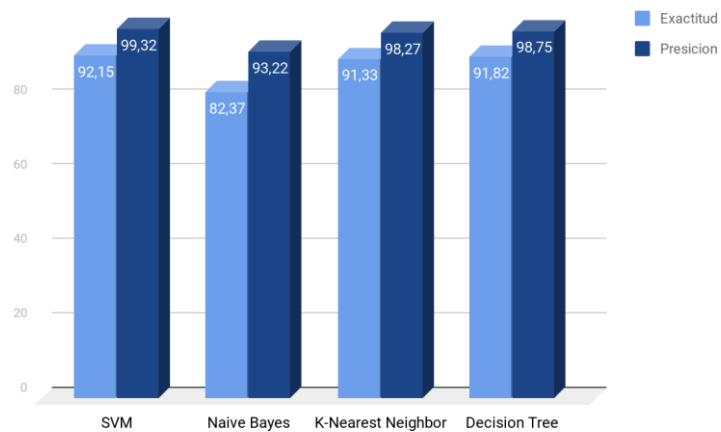
El artículo *“Leveraging Human and Machine Learning for Crisis Mapping during Disasters using Social Media”*, utiliza dos modelos para predecir la gravedad de los desastres naturales mediante la data recopilada de la API de Twitter, evaluó el desempeño de los modelos mediante su exactitud, nivel de confianza y exhaustividad, en la Figura 4 se observa que el modelo SSVP tiene mayor nivel de exactitud con un 91.01%, confianza con 91% y exhaustividad con 95%.

FIGURA 4. MEDICION DE MODELOS



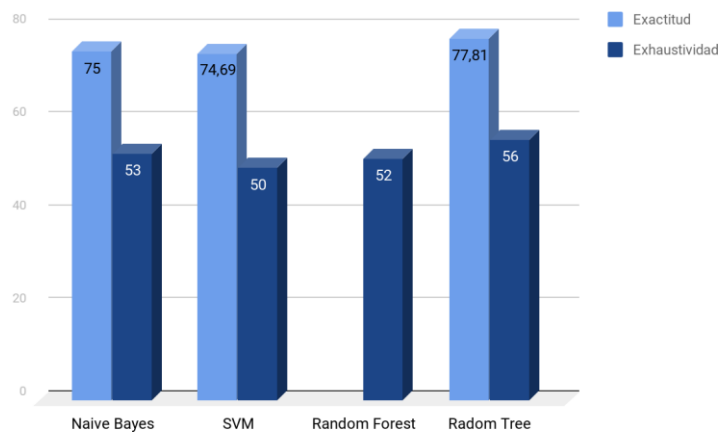
El artículo *“Predicting the sentiment of SaaS online reviews using supervised machine learning techniques”* analizó los comentarios y opiniones de los consumidores de productos de software como servicio (SaaS). Se evaluó los modelos bajo índices de categorización de texto: exactitud y precisión. Sus resultados indican que SVM tiene la mejor exactitud de 92,15% y precisión de 99,32%.

FIGURA 5. RESULTADOS DE PRECISIÓN Y EXACTITUD



El artículo *“Automated Detection of Hate Speech towards Women en Twitter”*, detecta el odio hacia las mujeres en la plataforma de Twitter, selecciono los siguientes modelos: Naive Bayes, SVM, Random Forest, Random Tree evaluados por la precisión y exactitud.

FIGURA 6. COMPARACIÓN DE EVALUACIÓN DE MODELOS



4.6.2 Pregunta de Investigación 2 (PI-2)

¿Qué funciones cumplen los modelos de Aprendizaje supervisado como apoyo a la toma de decisiones organizacional basados en datos de redes sociales?

En las investigaciones se encontró que los modelos de AS cumplen diferentes funciones que permiten una toma de decisiones organizacional más eficiente, clara y objetiva, estas son:

TABLA 15. FUNCIONES QUE REALIZAN LOS MODELOS

Funciones/Modelos	NB	SVM	NN	DT	KNN	Regresión
Identificación y predicción de peligros cibernético	x	x	x	x	x	
Identificación de las opiniones de la comunidad frente a temas políticos y predicción de resultados electorales	x	x		x	x	x
Identificación de la percepción de la comunidad frente a un producto o servicio de una empresa	x	x		x	x	
Identificación de perfiles de personalidad		x	x	x		x
Construcción de sistemas de Recomendación inteligente		x	x	x	x	x
Predicción de la tasa			x			

de ocupación de hoteles				
Predicción de deserción de atletas universitarios	x		x	x
Predicción del mercado de valores	x		x	
Identificación de las opiniones de la comunidad acerca de las vacunas		x		
Predicción de los ingresos de taquilla de películas				x
Identificación y predicción de desastres naturales	x	x	x	

Como se ha detallado en la tabla 15 los modelos permiten : **La identificación y predicción de peligros cibernético** mediante la clasificación y predicción de acoso cibernético, odio hacia un género humano y bullying cibernético. De manera que las organizaciones que buscan combatir casos de acoso, bullying u otro tipo de agresión tomen decisiones sobre las medidas preventivas y planes de concientización a realizar, así mismo ayuda a los investigadores que persiguen la tarea de detección de ciberacoso [1][4][8][25]. **Identificación de las opiniones de la comunidad frente a temas políticos y predicción de resultados electorales**, es un tema que tomó importancia al ser utilizada por el Ex Presidente de los Estados Unidos Barack Obama que ayudó a predecir los resultados de su campaña electoral y conocer las debilidades, fortalezas y la necesidad de la comunidad con respecto a su partido político, de modo que se tomen decisiones objetivas como el diseño de campañas de marketing [7][12][21][30][31]. **Identificación de la percepción de la comunidad frente a un producto o servicio de una empresa**, permiten identificar opiniones positivas, negativas y neutras sobre un producto o servicio. De manera que las empresas teniendo esta información puedan planificar estrategias de marketing según los comentarios y evaluaciones de sus usuarios. Hay muchas compañías conocidas como Amazon, Google, Twitter y TripAdvisor que han acumulado una gran colección de opiniones y han analizado sus servicios mediante algoritmos inteligentes [3][6][15][18][24][28][29]. **Identificación de perfiles de personalidad**, es un estudio que permite que las organizaciones puedan ofrecer producto y/o servicios de acuerdo a un sector de personas con gustos y preferencias similares. Así también ayuda a predecir trastornos de personalidad e implementar medidas preventivas para aquellas personalidades peligrosas [32]. **Construcción de sistemas de Recomendación inteligente**, proponen recomendaciones especializadas en base a los gustos, comportamiento y preferencias de cada usuario. Un ejemplo son los sistemas de recomendación de viajes de gestión para vacaciones de turistas en base a datos de TripAdvisor, o sistemas de recomendación de aplicaciones para uso empresarial [14][27]. **Predicción de la tasa de ocupación de hoteles** que permiten a los servicios de hostelería planificar planes de servicio y tomar decisiones con respecto a los servicios y productos a ofrecer [26]. **Predicción de deserción de atletas universitarios** permiten a los reclutadores de deportistas y a entrenadores tomar medidas preventivas para evitar la deserción de sus estudiantes, de manera que su inversión no sea desaprovechada [23]. **Predicción del mercado de valores**, esto es de vital importancia para los acciones para saber si deben de invertir en la compra de acciones de una empresa, de manera que se debe conocer la relación cliente-empresa[10].

Identificación de las opiniones de la comunidad acerca de las vacunas, no se realizan muchas investigaciones sobre este tema, sin embargo cabe resaltar la importancia de identificar patrones de rechazo en la personas frente a sistemas de vacunación de prevención de manera que las organizaciones de salud tomen decisiones para disminuir las muertes de enfermedades que pueden ser contrarrestadas por medio de vacunas [11]. **Predicción de los ingresos de taquilla de películas** esto permite a las empresas del rubro de entretenimiento conocer las películas de mayor frecuencia de visualización para diseñar promociones objetivas y aumentar sus ingresos en ventas[17]. Y por último la **identificación y predicción de desastres naturales** permiten a las organizaciones de anti desastres conocer la propensión de desastres, los riesgos producidos y la necesidad de la personas en situaciones de desastre, de manera que se tenga una visión más precisa para tomar medidas de contingencia y brindar capacitaciones a la población sobre las acciones que se deben realizar [2][20].

4.6.3 Pregunta de Investigación 3 (PI-3)

¿Qué herramientas se utilizaron en los modelos de aprendizaje supervisado encontrados?

En las investigaciones se identifico herramientas que ayudaron a los modelos de aprendizaje supervisado con la su implementación:

- **Weka:** Es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, agrupamiento, minería de reglas de asociación y visualización.
- **Rapidminer:** Es un programa informático para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. Se usa en investigación, educación, capacitación, creación rápida de prototipos y en aplicaciones empresariales.
- **Scikit-Learn:** Es una biblioteca para aprendizaje de máquina de software libre para el lenguaje de programación Python. Incluye varios algoritmos de clasificación, regresión y análisis de grupos entre los cuales están Support Vector Machine, Forest Random, Gradient boosting, K-means y DBSCAN. Está diseñada para interoperar con las bibliotecas numéricas y científicas NumPy y SciPy.

- **Apache Spark:** Se puede considerar un sistema de computación en clúster de propósito general y orientado a la velocidad. Proporciona APIs en Java, Scala, Python y R. También proporciona un motor optimizado que soporta la ejecución de grafos en general. También soporta un conjunto extenso y rico de herramientas de alto nivel entre las que se incluyen Spark SQL, MLlib para implementar machine learning, GraphX para el procesamiento de grafos y Spark Streaming.
- **Nltk:** El kit de herramientas de lenguaje natural, o más comúnmente NLTK, es un conjunto de bibliotecas y programas para el procesamiento del lenguaje natural (PLN) simbólico y estadísticos para el lenguaje de programación Python. NLTK incluye demostraciones gráficas y datos de muestra. NLTK está destinado a apoyar la investigación y la enseñanza en PLN o áreas muy relacionadas, que incluyen la lingüística empírica, las ciencias cognitivas, la inteligencia artificial, la recuperación de información, y el aprendizaje de la máquina.
- **Matlab:** Entre sus prestaciones básicas se hallan la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario y la comunicación con programas en otros lenguajes y con otros dispositivos hardware.
- **QDA Miner:** Es un software de análisis de datos cualitativos y métodos mixtos desarrollado por Provalis Research. El programa fue diseñado para ayudar a los investigadores a gestionar, codificar y analizar datos cualitativos.
- **WordStat:** Es un software de análisis de contenido y minería de texto. El software se utiliza principalmente para inteligencia empresarial y análisis competitivo de sitios web, análisis de sentimientos, análisis de contenido de preguntas abiertas, extracción de temas de datos de redes sociales.
- **Knime:** Es una plataforma de minería de datos que permite el desarrollo de modelos en un entorno visual. Está construido bajo la plataforma Eclipse.
- **LibSVM:** Es una biblioteca popular de aprendizaje automático de código abierto. El código de aprendizaje SVM de la biblioteca a menudo se reutiliza en otros kits de herramientas de aprendizaje automático de código abierto, incluidos GATE , KNIME , Orange y scikit-learn . Existen enlaces y puertos para lenguajes de programación como Java , MATLAB , R y Python.
- **Statistica 13:** Es un paquete estadístico usado en investigación, minería de datos y en el ámbito empresarial. Lo creó StatSoft, empresa que lo desarrolla y mantiene.
- **Mathematica:** Es un programa utilizado en áreas científicas, de ingeniería, matemática y áreas computacionales. Comúnmente considerado como un sistema de álgebra computacional, Mathematica es también un poderoso lenguaje de programación de propósito general.

4.7 Amenazas de la validez

A continuación, se procede con la discusión de las amenazas para la validez de la presente revisión. Para la ejecución de la cadena de búsqueda, se consideraron 4 bases de datos digitales y librerías indexadas, sin embargo, al implementar las cadenas de búsquedas, se encontraron artículos que aun con todos los filtros se visualiza una gran cantidad, por lo que se optó por modificar en el buscador las cadenas de manera que reduzca el tamaño para ejecutar los criterios de inclusión y exclusión. La selección de los estudios primarios, a pesar de haber seguido criterios de inclusión y exclusión definidos claramente, puede haber sido afectada por la poca información presentada en los resúmenes de los artículos, ya que en algunos casos esta no era presentada de manera clara. Se provee toda la información necesaria para poder realizar la réplica del presente estudio. Todas las librerías digitales y artículos utilizados se encuentran propiamente referenciados y disponibles en internet.

5 Conclusiones

En este estudio se presentan los resultados de una revisión sistemática realizada a 32 artículos encontrados en 4 librerías digitales y bases de datos indexadas de gran relevancia en el ámbito científico y académico. Asimismo, se presenta dentro del análisis bibliométrico la clasificación de los estudios por año de publicación, donde se puede observar un incremento en el número de estudios publicados desde el año 2015 a 2017, lo que prueba que existe un interés en la utilización de modelos de aprendizaje supervisado como apoyo a la toma de decisiones en las organizaciones basados en datos de redes sociales. La concentración de artículos en los dominios de publicación:

Data and Information, Science and Information, Computer Science, Information and Software Technology, en la destaca en la “*International Conference on Advances in Social Networks Analysis and Mining*”, “*Computers in Human Behavior*”, “*Tourism Management*”, también se encontró que el 50% de las publicaciones son provenientes de revistas de investigación.

Con respecto a las preguntas de investigación planteadas, se obtuvieron 6 modelos de aprendizaje supervisado, que son: El modelo de Support Vector Machine, el modelo de Decision Tree, el modelo de Neural Network, el modelo de Regression: Logistic y Lineal, el modelo de Naive Bayes y el modelo K-Nearest Neighbors. También se ha encontrado que Twitter es la red social más utilizadas para la extracción de datos y la aplicación de algoritmos inteligentes en base a las cadenas de datos brindados por las mismas. Y finalmente se encontró que los modelos mencionados aportan a la toma de decisiones organizacional brindando información valiosa y procesada, y mecanismos para la predicción, clasificación o agrupación de datos que permiten la Identificación y predicción de peligros cibernético, Identificación de las opiniones de la comunidad frente a temas políticos y predicción de resultados electorales, Identificación de la percepción de la comunidad frente a un producto o servicio de una empresa, identificación de perfiles de personalidad, construcción de sistemas de Recomendación inteligente, predicción de la tasa de ocupación de

hoteles, predicción de deserción de atletas universitarios, predicción del mercado de valores, identificación de las opiniones de la comunidad acerca de las vacunas, predicción de los ingresos de taquilla de películas, identificación y predicción de desastres naturales. De manera que esta información procesada tenga un valor comercial significativo y sea un base para el proceso de toma de decisiones. Es así como el presente estudio ha llevado a proponer como trabajo futuro la implementación de un modelo de aprendizaje para el análisis de sentimientos sobre las opiniones de los estudiantes con respecto a la realización de cultos que realiza la Universidad Peruana Unión en base a los datos brindados en redes sociales principalmente de las publicaciones de Facebook de la universidad.

Referencias

- [1] C. G. Zazo and S. R. Crespo, "Las Redes Sociales Como Impulsoras del Aprendizaje Colaborativo en Educación."
- [2] C. Isabel and H. Malaquina, "Las Redes Sociales , ¿ Han Cambiado La Forma De Abordar Los Procesos De Reclutamiento Y Selección ? Have Social Networks Changed the Way To Approach the Processes of Selection and Recruitment ?," *Rev. Iberoam. ciencias Empres. y Econ.*, vol. 7, 2016.
- [3] J. Freire, "Social networks: Organizational models or digital services?," *Prof. la Inf.*, vol. 17, no. 6, pp. 585–588, 2008.
- [4] S. Otero, "Aportes De La Ingeniería Para El Desarrollo Regional," *Utm*, no. November 2017, 2017.
- [5] S. M. Tenzer and O. Ferro, "Redes Sociales Virtuales: personas, sociedad y empresa Contenido," pp. 1–18, 2009.
- [6] H. Hern, "Aplicativo Web Para Localizar Datos de Contactos Personales Utilizando Técnicas de Minería Web y Redes Sociales," pp. 1–6, 2014.
- [7] J. Freire, "Social networks: Organizational models or digital services?," *Prof. la Inf.*, vol. 17, no. 6, pp. 585–588, 2008.
- [8] C. E. López Zambrano, "Implementación de un sistema de chatbot para la atención de consultas de información a través de las redes sociales de las Carreras de Ingeniería en Sistemas Computacionales y Computación de la Universidad Católica de Santiago de Guayaquil.," p. 148, 2018.
- [9] A. Baj-Rogowska, "Sentiment analysis of Facebook posts: The Uber case," *2017 IEEE 8th Int. Conf. Intell. Comput. Inf. Syst. ICICIS 2017*, vol. 2018-January, no. Iccis, pp. 391–395, 2018.

- [10] C. G. Rubio, "Clasificación automática de texto para el seguimiento de campañas electorales en redes sociales," 2015.
- [11] M. del R. Martínez Torres, D. Gutiérrez Reina, S. L. Toral, and F. Barrero, "Metodologías de análisis de los big data en las plataformas educativas," pp. 79–83, 2014.
- [12] M. Leyva Vázquez, R. Escobar Jara, C. Espín Riofrio, and K. Pérez Teruel, "Facebook como herramienta para el aprendizaje colaborativo de la inteligencia artificial," *Didasc@lia Didáctica y Educ.*, vol. 9, no. 1, pp. 27–36, 2018.
- [13] P. Rocha, "Conceptos básicos, modelos y simulación," *Model. Digit. del Terreno*, p. 122, 1994.
- [14] C. Manterola, P. Astudillo, E. Arias, and N. Claros, "Revisiones sistemáticas de la literatura. Qué se debe saber acerca de ellas," *Cir. Esp.*, vol. 91, no. 3, pp. 149–155, 2013.
- [15] S. Kitchenham, B. and Charters, "Guidelines for performing systematic literature reviews in software engineering," *Tech. report, Ver. 2.3 EBSE Tech. Report. EBSE*, 2007.
- [16] C. M. D. C. Santos, C. A. D. M. Pimenta, and M. R. C. Nobre, "A estratégia PICO para a construção da pergunta de pesquisa e busca de evidências," *Rev. Lat. Am. Enfermagem*, vol. 15, no. 3, pp. 508–511, 2007.
- [17] J. Popay, L. Arai, and N. Britten, "Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme Information needs of adolescents with Adolescent Idiopathic Scoliosis: a multifaceted study View Project The SENS project View project," no. May 2014, 2006.
- [18] D. Rouhani, M. N. Mahrin, F. Nikpay, R. B. Ahmad y P. Nikfard, "Asystematic literature review on Enterprise Architecture Implementation Methodologie", *Information and Software Technology*, pp. 1-20, 2015.
- [19] M. De, T. Doctoral, P. Por, F. Fernando De La, and R. Troyano, "UNIVERSIDAD DE SEVILLA Sistemas de Inteligencia Web: Análisis de redes sociales," 2012.
- [20] A. Kuz, M. Falco, L. Nahuel, and R. Giandini, "Integrando Redes Sociales y Técnicas de Inteligencia Artificial en Entornos Educativos," *Rev. Q*, vol. 10, no. 19, pp. 1–9, 2015.
- [21] P. Pazmiño, "UNIVERSIDAD POLITÉCNICA SALESIANA SEDE QUITO Tesis previa a la obtención del Título de: Licenciada en comunicación social TÍTULO O TEMA El impacto de las redes sociales y el internet en la formación de los jóvenes de la Universidad Politécnica Salesiana," *Impacto en la Redes Soc.*, pp. 1–139, 2010.

[22] S. Kemp, "Global Digital Report 2019 - We Are Social", We Are Social, 2019. [Online]. Available: <https://wearesocial.com/global-digital-report-2019>. [Accessed: 10- Nov- 2019].

[23]
"http://biblioteca.clacso.edu.ar/Argentina/unlar/20171117044954/pdf_1512.pdf." 2015.

[24] O. González M.R. y Carrión, "8. Aplicación de las redes sociales en la empresa – Francisco Galán – 2011," 2012.

[25] D. I. Candia Oviedo, "Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático," *Univ. Nac. San Antonio Abad del Cusco*, 2019.

[26] J. Lozano Silva, "Aprendizaje supervisado eficiente para el análisis de datos geoespaciales a gran escala," no. c, p. 1, 2015.

[27] E. H. K. C, "ONLINE SOCIAL NETWORK ANALYSIS USING MACHINE LEARNING TECHNIQUES," vol. 4, no. 4, pp. 25–40, 2017.

[28] A. Flores, "El impacto de las redes sociales como herramienta didáctica en el aprendizaje de los estudiantes de la carrera de Computación e Informática del Instituto Superior de Educación Público Chíncha , 2015," p. 157, 2017.

[29] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts," *Math. Comput. Appl.*, vol. 23, no. 1, p. 11, 2018.

[30] J. A. Naser, "Neural networks - a brief introduction," *Proc. Am. Power Conf.*, vol. 53, no. pt 2, pp. 943–945, 1991.

[31] P. Langley and J. G. Carbonell, *Approaches to machine learning*, vol. 35, no. 5. 1984.

[32] G. De la Calle Velasco, "Modelo basado en técnicas de procesamiento de lenguaje natural para extraer y anotar información de publicaciones científicas," p. 191, 2014.

[33] Docs.aws.amazon.com. (2019). Entrenamiento de modelos de ML - Amazon Machine Learning. [online] Available at: https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/training-ml-models.html [Accessed 12 Nov. 2019].

APÉNDICE

A. Artículos seleccionados

ID	Biblioteca	Título	Autor	Año	Tipo	Dominio
1	IEEE Xplore Digital Library	Weakly Supervised Cyberbullying Detection Using Co-Trained Ensembles of Embedding Models	Elaheh Raisi, Bert Huang	2018	Conference	Computer Science
2	IEEE Xplore Digital Library	Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media	Bhuvaneswari Anbalagan, C. Valliammai	2016	Conference	Data and Information
3	IEEE Xplore Digital Library	Predicting the sentiment of SaaS online reviews using supervised machine learning techniques	Asma Musabah Alkalbani, Ahmed Mohamed Ghamry, Farookh Khadeer Hussain, Omar Khadeer Hussain	2016	Conference	Information and Software Technology
4	IEEE Xplore Digital Library	Automated Detection of Hate Speech towards Woman on Twitter	Havvanur Şahi, Yasemin Kılıç, Rahime Belen Sağlam	2018	Conference	Science and Information
5	IEEE Xplore Digital Library	Sentiment Analysis of Facebook Posts: the Uber case	Anna Baj-Rogowska	2017	Conference	Data and Information

6	IEEE Xplore Digital Library	Analysing Cloud Services Reviews Using Opining Mining	Asma Musabah Alkalbani, Lekhaben Gadhvi, Bhaumik Patel, Farookh Khadeer Hussain, Ahmed Mohamed Ghamry, Omar Khadeer Hussain	2015	Conference	Computer Science
7	IEEE Xplore Digital Library	Election result prediction using Twitter sentiment analysis	Jyoti Ramteke, Samarth Shah, Darshan Godhia, Aadil Shaikh	2016	Conference	Data and Information
8	IEEE Xplore Digital Library	A Bag-of-Phonetic-Codes Model for Cyber-Bullying Detection in Twitter	Ankita Shekhar, M. Venkatesan	2018	Conference	Data and Information
9	IEEE Xplore Digital Library	Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction	Moh. Nasrul Aziz, Ari Firmanto, A. Miftah Fajrin, R. V. Hari Ginardi	2018	Conference	Data and Information

10	IEEE Xplore Digital Library	Stock price prediction using linear regression based on sentiment analysis	Yahya Eru Cakra, Bayu Distiawan Trisedya	2015	Conference	Data and Information
11	ACM Digital Library	Examining Online Vaccination Discussion and Communities in Twitter	Xiaoyi Yuan, Andrew T. Crooks	2018	Conference	Data and Information
12	ACM Digital Library	Detection of Political Manipulation in Online Communities through Measures of Effort and Collaboration	Brian D. Davison, Mariann Winslett	2015	Academy	Science and Information
13	ACM Digital Library	A Supervised Learning Method for Prediction Citation Count of Scientists in Citation Networks	Jana Diesner, Elena Ferrari, Guandong Xu	2017	Conference	Science and Information
14	ACM Digital Library	Hashtag Recommendation for Enterprise Applications	Dhruv Mahajan, Vishwajit Kolathur, Chetan Bansal, Suresh Parthasarathy, S. Sundararajan, Sathiya Keerthi, Johannes	2016	Conference	Science and Information

			Gehrke			
15	ACM Digital Library	An Apache Spark implementation for graph-based hashtag sentiment classification on Twitter	Elias Dritsas, Konstantinos Giotopoulos, Ioannis E. Livieris, Leonidas Theodorakopoulos	2018	Conference	Science and Information
16	ACM Digital Library	Learning to Infer Competitive Relationships in Heterogeneous Networks	Yang Yang, Jie Tang, Juanzi Li	2018	Journal	Data and Information
17	ACM Digital Library	Social Media Mining: Prediction of Box Office Revenue	Deepankar Choudhery, Carson K. Leung	2017	Conference	Data and Information
18	ACM Digital Library	Sentiment analysis at sentence level for heterogeneous datasets	Jawad Khan, Byeong Soo Jeong, Young-Koo Lee, Aftab Alam	2016	Conference	Data and Information
19	Science Direct	Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter	Hamid Shirdastan, Michel Laroche, Marie-	2017	Journal	Data and Information

Odile Richard						
20	Science Direct	Usage and analysis of Twitter during 2015 Chennai flood towards disaster management	Meera.R. Nair , G.R.Ramya , P.Bagavathi Sivakumar	2017	Conference	Data and Information
21	Science Direct	Pre Processing of Twitter's Data for Opinion Mining in Political Context	Ratab Gull , Umar Shoaib , Saba Rasheed, Washma Abid , Beenish Zahoor	2016	Conference	Data and Information
22	Science Direct	Online reputation measurement of companies based on user-generated content in online social networks	Hossein Shad Manaman, Shahram Jamali , Abolfazl AleAhmad	2015	Journal	Science and Information
23	Science Direct	The turf is always greener: Predicting decommitments in college football recruiting using Twitter data	Kristina Gavin Bigsby , Jeffrey W. Ohlmann , Kang Zhao	2019	Journal	Science and Information
24	Science Direct	A framework for big data analytics in commercial social networks: A	Erick Kauffman , Jesús	2019	Journal	Data and Information

		case study on sentiment analysis and fake review detection for marketing decision-making	Peral , David Gil,, Antonio Ferrández , Ricardo Sellers, Higinio Mora			
25	Science Direct	Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network	Mohammed Ali Al-garadi , Kasturi Dewi Varathan , Sri Devi Ravana	2016	Journal	Data and Information
26	Science Direct	Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities	Enrique Bigné , Enrique Oltra , Luisa Andreu	2019	Journal	Science and Information
27	Science Direct	Facilitating tourists' decision making through open data analyses: A novel recommender system	Eleonora Pantano , Constantinos-Vasilios Priporas, Nikolaos Stylos, Charles Dennis	2019	Journal	Data and Information
28	Springer Link	An exploratory study of Twitter messages about software applications	Emitza Guzman, Rana Alkadhi, Norbert	2017	Journal	Science and Information

Seyff						
29	Springer Link	DERIV: distributed brand perception tracking framework	Manu Shukla, Raimundo Dos Santos, Andrew Fong, Chang-Tien Lu	2017	Journal	Science and Information
30	Springer Link	Hybrid N-gram model using Naïve Bayes for classification of political sentiments on Twitter	Jamilu Awwalu, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub	2019	Journal	Data and Information
31	Springer Link	Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning	Ronaldo Cristiano Prati, Elias Said-Hung	2019	Journal	Science and Information
32	Springer Link	Personality Prediction System from Facebook Users	Tommy Tandra, Hendro, Derwin Suharton, Rini Wongso, and Yen Lina Prasetio	2017	Conference	Data and Information

B. Formularios de extracción

criterio	Detalle	Relevancia
----------	---------	------------

Identificador	1	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Weakly Supervised Cyberbullying Detection Using Co-Trained Ensembles of Embedding Models	PB-2
Autores	Elaheh Raisi, Bert Huang	PB-2
Publicación	IEEE/ACM Digital Library International Conference on Advances in Social Networks Analysis and Mining (ASONAM)	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Recurrent Neural Network	PI-1
Funciones	Detección del bullying cibernético (Twitter, Instagram, Ask)	PI-2

Criterio	Detalle	Relevancia
Identificador	2	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media	PB-2
Autores	Bhuvanewari Anbalagan, C. Valliammai	PB-2
Publicación	IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW)	PB-3
Año de publicación	2016	PB-1

Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, SVM	PI-1
Funciones	Identificación y clasificación de alertas de desastres naturales tempranas de datos de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	3	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Predicting the sentiment of SaaS online reviews using supervised machine learning techniques	PB-2
Autores	Asma Musabah Alkalbani, Ahmed Mohamed Ghamry, Farookh Khadeer Hussain, Omar Khadeer Hussain	PB-2
Publicación	International Joint Conference on Neural Networks (IJCNN)	PB-3
Año de publicación	2016	PB-1
Tipo de publicación	Conference	PB-2
Modelos de AS	Support Vector Machine, Naive Bayes, k-Nearest Neighbors, Random Tree	PI-1
Funciones	Clasificación de análisis de sentimientos sobre productos lanzados línea	PI-2

Criterio	Detalle	Relevancia
Identificador	4	-
Fuente	IEEE Xplore Digital Library	PB-2

Título	Automated Detection of Hate Speech towards Woman on Twitter	PB-2
Autores	Havvanur Şahi, Yasemin Kılıç, Rahime Belen Sağlam	PB-2
Publicación	3rd International Conference on Computer Science and Engineering (UBMK)	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Revistas	PB-2
Modelos de AS	Support Vector Machine, Naive Bayes, Random Tree	PI-1
Funciones	Detección y clasificación automática del odio cibernético hacia las mujeres en Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	5	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Sentiment Analysis of Facebook Posts: the Uber case	PB-2
Autores	Anna Baj-Rogowska	PB-2
Publicación	The 8th IEEE International Conference on Intelligent Computing and Information Systems	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	k-Nearest Neighbors, Random Tree , Neural Network, Support Vector Machine, Naive Bayes, Linear Regression, No Linear Regression.	PI-1

Funciones	Analizar las opiniones en facebook sobre Uber	PI-2
-----------	---	------

Criterio	Detalle	Relevancia
Identificador	6	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Analysing Cloud Services Reviews Using Opining Mining	PB-2
Autores	Asma Musabah Alkalbani, Lekhaben Gadhvi, Bhaumik Patel, Farookh Khadeer Hussain, Ahmed Mohamed Ghamry, Omar Khadeer Hussain	PB-2
Publicación	IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)	PB-3
Año de publicación	2015	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	k-Nearest Neighbors, Naive Bayes, Random Tree, Random Forest	PI-1
Funciones	Predicción de sentimientos de usuarios hacia un producto en línea datos de Amazon, Google, Twitter, y TripAdvisor	PI-2

Criterio	Detalle	Relevancia
Identificador	7	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Election result prediction using Twitter sentiment analysis	PB-2
Autores	Jyoti Ramteke, Samarth Shah, Darshan	PB-2

Godhia, Aadil Shaikh		
Publicación	International Conference on Inventive Computation Technologies (ICICT)	PB-3
Año de publicación	2016	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, Support Vector Machine	PI-1
Funciones	Predicción de los resultados de las elecciones mediante el análisis de sentimientos de base de datos de Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	8	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	A Bag-of-Phonetic-Codes Model for Cyber-Bullying Detection in Twitter	PB-2
Autores	Ankita Shekhar, M. Venkatesan	PB-2
Publicación	International Conference on Current Trends towards Converging Technologies (ICCTCT)	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, Support Vector Machine	PI-1
Funciones	Detección de acoso cibernético en la plataforma de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	9	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction	PB-2
Autores	Moh. Nasrul Aziz, Ari Firmanto, A. Miftah Fajrin, R. V. Hari Ginardi	PB-2
Publicación	5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Análisis de Sentimientos: Detectar la satisfacción de los servicios gubernamentales en base a la plataforma de Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	10	-
Fuente	IEEE Xplore Digital Library	PB-2
Título	Stock price prediction using linear regression based on sentiment analysis	PB-2
Autores	Yahya Eru Cakra, Bayu Distiawan Trisedya	PB-2
Publicación	International Conference on Advanced Computer Science and Information Systems (ICACSIS)	PB-3
Año de publicación	2015	PB-1

Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, Forest Random	PI-1
Funciones	Predicción del mercado de valores utilizando análisis de sentimientos en Indonesia.	PI-2

Criterio	Detalle	Relevancia
Identificador	11	-
Fuente	ACM Digital Library	PB-2
Título	Examining Online Vaccination Discussion and Communities in Twitter	PB-2
Autores	Xiaoyi Yuan, Andrew T. Crooks	PB-2
Publicación	9th International Conference on Social Media and Society	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Determinar la influencia de las vacunas en una comunidad en base a la plataforma de Twitter: Análisis de Sentimientos.	PI-2

Criterio	Detalle	Relevancia
Identificador	12	-
Fuente	ACM Digital Library	PB-2

Título	Detection of Political Manipulation in Online Communities through Measures of Effort and Collaboration	PB-2
Autores	Brian D. Davison, Marianne Winslett	PB-2
Publicación	ACM Transactions on the Web (TWEB)	PB-3
Año de publicación	2015	PB-1
Tipo de publicación	Academica	PB-2
Modelos de AS	Support Vector Machine, Random Forest	PI-1
Funciones	Detección de la manipulación de política en una elección en base a datos de Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	13	-
Fuente	ACM Digital Library	PB-2
Título	A Supervised Learning Method for Prediction Citation Count of Scientists in Citation Networks	PB-2
Autores	Jana Diesner, Elena Ferrari, Guandong Xu	PB-2
Publicación	International Conference on Advances in Social Networks Analysis and Mining	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	k-Nearest Neighbors, Logistic Regression, Neural Network	PI-1
Funciones	Predicción del recuento de citas futuras de un científico determinado	PI-2

Criterio	Detalle	Relevancia
Identificador	14	-
Fuente	ACM Digital Library	PB-2
Título	Hashtag Recommendation for Enterprise Applications	PB-2
Autores	Dhruv Mahajan, Vishwajit Kolathur, Chetan Bansal, Suresh Parthasarathy, S. Sundararajan, Sathiya Keerthi, Johannes Gehrke	PB-2
Publicación	25th ACM International on Conference on Information and Knowledge Management	PB-3
Año de publicación	2016	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	KNN, Random Forest, Logistic Regression	PI-1
Funciones	Detectar auto recomendación de aplicaciones empresariales	PI-2

Criterio	Detalle	Relevancia
Identificador	15	-
Fuente	ACM Digital Library	PB-2
Título	An Apache Spark implementation for graph-based hashtag sentiment classification on Twitter	PB-2
Autores	Elias Dritsas, Konstantinos Giotopoulos, Ioannis E. Livieris, Leonidas Theodorakopoulos	PB-2
Publicación	22nd Pan-Hellenic Conference on Informatics	PB-3

Año de publicación	2018	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Análisis de sentimientos sobre productos y servicios basados en datos de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	16	-
Fuente	ACM Digital Library	PB-2
Título	Learning to Infer Competitive Relationships in Heterogeneous Networks	PB-2
Autores	Yang Yang, Jie Tang, Juanzi Li	PB-2
Publicación	ACM Transactions on Knowledge Discovery from Data (TKDD)	PB-3
Año de publicación	2018	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Linear Regression, Support Vector Machine	PI-1
Funciones	Predicción y clasificación de relaciones competitivas en empresas usando Twitter y registros patentes.	PI-2

Criterio	Detalle	Relevancia
Identificador	17	-

Fuente	ACM Digital Library	PB-2
Título	Social Media Mining: Prediction of Box Office Revenue	PB-2
Autores	Deepankar Choudhery, Carson K. Leung	PB-2
Publicación	21st International Database Engineering & Applications Symposium	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Regression Linear	PI-1
Funciones	Predecir los ingresos de taquilla de las películas mediante la extracción de características como los tweets y sus sentimientos.	PI-2

Criterio	Detalle	Relevancia
Identificador	18	-
Fuente	ACM Digital Library	PB-2
Título	Sentiment analysis at sentence level for heterogeneous datasets	PB-2
Autores	Jawad Khan, Byeong Soo Jeong, Young-Koo Lee, Aftab Alam	PB-2
Publicación	Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory	PB-3
Año de publicación	2015	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	k-Nearest Neighbors	PI-1

Funciones	Análisis de sentimientos sobre la satisfacción de un servicio y producto basado en información de Facebook y Twitter	PI-2
-----------	--	------

Criterio	Detalle	Relevancia
Identificador	19	-
Fuente	Science Direct	PB-2
Título	Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter	PB-2
Autores	Hamid Shirdastian , Michel Laroche , Marie-Odile Richard	PB-2
Publicación	International Journal of Information Management	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Análisis del sentimiento de autenticidad de marca Starbucks en Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	20	-
Fuente	Science Direct	PB-2
Título	Usage and analysis of Twitter during 2015 Chennai flood towards disaster management	PB-2
Autores	Meera.R.Nair , G.R.Ramya , P.Bagavathi Sivakumar	PB-2

Publicación	7th International Conference on Advances in Computing & Communications	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, Random Forest, Random Tree	PI-1
Funciones	Identificar áreas afectadas y la opinión de la gente en las inundaciones de Chennai 2015 a través de Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	21	-
Fuente	Science Direct	PB-2
Título	Pre Processing of Twitter's Data for Opinion Mining in Political Context	PB-2
Autores	Ratab Gull , Umar Shoaib , Saba Rasheed, Washma Abid , Beenish Zahoor	PB-2
Publicación	20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems	PB-3
Año de publicación	2016	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Naive Bayes, Support Vector Machine	PI-1
Funciones	Análisis de sentimientos de los usuarios de Twitter sobre contextos políticos.	PI-2

Criterio	Detalle	Relevancia
Identificador	22	-
Fuente	Science Direct	PB-2
Título	Online reputation measurement of companies based on user-generated content in online social networks	PB-2
Autores	Hossein Shad Manaman, Shahram Jamali , Abolfazl AleAhmad	PB-2
Publicación	Computers in Human Behavior	PB-3
Año de publicación	2015	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Neural Network, Naive Bayes	PI-1
Funciones	Medir la reputación de una empresa determinada en base a las opiniones de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	23	-
Fuente	Science Direct	PB-2
Título	The turf is always greener: Predicting decommitments in college football recruiting using Twitter data	PB-2
Autores	Kristina Gavin Bigsby , Jeffrey W. Ohlmann, Kang Zhao	PB-2
Publicación	Journal of Biomedical Informatics: X	PB-3
Año de publicación	2019	PB-1
Tipo de	Journal	PB-2

publicación		
Modelos de AS	Logistic Regression, Naive Bayes, Forest Random, Random Tree	PI-1
Funciones	Predecir la deserción de atletas universitarios según datos de Twitter.	PI-2

Criterio	Detalle	Relevancia
Identificador	24	-
Fuente	Science Direct	PB-2
Título	A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making	PB-2
Autores	Erick Kauffmann , Jesús Peral , David Gil,, Antonio Ferrández , Ricardo Sellers, Higinio Mora	PB-2
Publicación	Industrial Marketing Management	PB-3
Año de publicación	2019	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Análisis de sentimientos para analizar las revisiones en línea de productos de Amazon.	PI-2

Criterio	Detalle	Relevancia
Identificador	25	-
Fuente	Science Direct	PB-2

Título	Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network	PB-2
Autores	Mohammed Ali Al-garadi , Kasturi Dewi Varathan, Sri Devi Ravana	PB-2
Publicación	Computers in Human Behavior	PB-3
Año de publicación	2016	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Naive Bayes, Support Vector Machine, k-Nearest Neighbors , Forest Random	PI-1
Funciones	Detectar cyberbullying en base a datos de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	26	-
Fuente	Science Direct	PB-2
Título	Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities	PB-2
Autores	Enrique Bigné , Enrique Oltra , Luisa Andreu	PB-2
Publicación	Tourism Management	PB-3
Año de publicación	2019	PB-1
Tipo de publicación	Conferencia	PB-2
Modelos de AS	Neural Network	PI-1
Funciones	Predecir la tasa de ocupación de los hoteles según eventos, atracciones, socialización y marketing de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	27	-
Fuente	Science Direct	PB-2
Título	Facilitating tourists' decision making through open data analyses: A novel recommender system	PB-2
Autores	Eleonora Pantano , Constantinos-Vasilios Priporas, Nikolaos Stylos, Charles Dennis	PB-2
Publicación	Tourism Management Perspectives	PB-3
Año de publicación	2019	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Neural Network, k-Nearest Neighbors, Support Vector Machine	PI-1
Funciones	Construir sistemas de recomendación de viajes de destino para vacaciones a turistas en base a datos de TripAdvisor	PI-2

Criterio	Detalle	Relevancia
Identificador	28	-
Fuente	Springer Link	PB-2
Título	An exploratory study of Twitter messages about software applications	PB-2
Autores	Emitza Guzman, Rana Alkadhi, Norbert Seyff	PB-2
Publicación	Requirements Engineering	PB-3
Año de publicación	2019	PB-1

Tipo de publicación	Journal	PB-2
Modelos de AS	Neural Bayes, Support Vector Machine, Forest Random	PI-1
Funciones	Clasificación automática de los tweets sobre aplicaciones de software	PI-2

Criterio	Detalle	Relevancia
Identificador	29	-
Fuente	Springer Link	PB-2
Título	DERIV: distributed brand perception tracking framework	PB-2
Autores	Manu Shukla, Raimundo Dos Santos, Andrew Fong, Chang-Tien Lu	PB-2
Publicación	Journal of Big Data	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Support Vector Machine	PI-1
Funciones	Determinar la percepción del usuario de una marca en base a datos de Twitter	PI-2

Criterio	Detalle	Relevancia
Identificador	30	-
Fuente	Springer Link	PB-2

Título	Hybrid N-gram model using Naïve Bayes for classification of political sentiments on Twitter	PB-2
Autores	Jamilu Awwalu, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub	PB-2
Publicación	Neural Computing and Applications	PB-3
Año de publicación	2019	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Naive Bayes	PI-1
Funciones	Analizar opiniones de los usuarios de Twitter sobre temas políticos	PI-2

Criterio	Detalle	Relevancia
Identificador	31	-
Fuente	Springer Link	PB-2
Título	Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning	PB-2
Autores	Ronaldo Cristiano Prati, Elias Said-Hung	PB-2
Publicación	AI & SOCIETY	PB-3
Año de publicación	2019	PB-1
Tipo de publicación	Journal	PB-2
Modelos de AS	Naive Bayes, Forest Random, Support Vector Machine, Logistic Regression	PI-1
Funciones	Identificar la ideología partidista durante las elecciones celebradas en España el 24 de mayo de	PI-2

2015 publicadas por los usuarios de Twitter

Criterio	Detalle	Relevancia
Identificador	32	-
Fuente	Science direct	PB-2
Título	Personality Prediction System from Facebook Users	PB-2
Autores	Tommy Tandra, Hendro, Derwin Suhartono, Rini Wongso, and Yen Lina Prasetio	PB-2
Publicación	2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI	PB-3
Año de publicación	2017	PB-1
Tipo de publicación	Conference	PB-2
Modelos de AS	Neural Network, Naive Bayes, Support Vector Machine	PI-1
Funciones	Predecir la personalidad del usuario de Facebook en base a 5 modelos de personalidad.	PI-2