

MINERÍA DE INTENCIONES A PARTIR DE UNA BASE DEL CONOCIMIENTO Y APRENDIZAJE AUTOMÁTICO SUPERVISADO

INTENTION MINING FROM KNOWLEDGE BASE AND SUPERVISED MACHINE LEARNING

Oswaldo E. Díaz-Rodríguez

Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Quito, (Ecuador).

Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Quito, (Ecuador).

E-mail: oswaldo.diaz@epn.edu.ec ORCID: <https://orcid.org/0000-0003-4001-1858>

María Gabriela Pérez Hernández

Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Quito, (Ecuador).

E-mail: maria.perez@epn.edu.ec ORCID: <https://orcid.org/0000-0001-9628-2767>

Recepción: 22/03/2021 **Aceptación:** 30/06/2021 **Publicación:** 29/09/2021

Citación sugerida:

Díaz-Rodríguez, O. E., y Pérez, M. G. (2021). Minería de intenciones a partir de una base del conocimiento y aprendizaje automático supervisado. *3C TIC. Cuadernos de desarrollo aplicados a las TIC*, 10(3), 65-101. <https://doi.org/10.17993/3ctic.2021.103.65-101>

RESUMEN

La falta de flexibilidad en los sistemas de información, ha llevado a los usuarios a utilizar sus propias estrategias, para realizar sus actividades diarias alineadas con los objetivos del negocio. De esta forma, los usuarios cumplen con sus funciones, mejoran su desempeño y ahorran recursos; especialmente tiempo. Se propone un método de minería de intenciones, que se basa en el aprendizaje automático supervisado, respaldado por una base del conocimiento y las reglas heurísticas de supervisión. Una base del conocimiento se estructura a partir de documentos comerciales multidisciplinares para cualquier negocio en general. De la aplicación del método propuesto a un negocio en particular (ventas), se obtienen las estrategias del usuario y a partir de éstas, se infieren sus intenciones en el desarrollo de sus actividades en el negocio de ventas, a través de un sistema de información. Este trabajo sugiere el desarrollo de sistemas de información flexibles y brinda a los gerentes de negocios una herramienta para identificar e implementar nuevas estrategias de negocio, basadas en las estrategias del usuario.

PALABRAS CLAVE

Sistema de información, Actividad del negocio, Estrategia de usuario, Base del conocimiento, Aprendizaje automático supervisado.

ABSTRACT

The lack of flexibility in information systems has led users to use their own strategies to carry out their daily activities in line with business objectives. In this way, users fulfill their functions, improve their performance and save resources; especially time. An intent mining method is proposed, which is based on supervised machine learning, supported by a knowledge base and heuristic business rules. A knowledge base is structured from multidisciplinary business documents for any business in general. From the application of the proposed method to a particular business (sales), the user's strategies are obtained and from these, their intentions in the development of their activities in the sales business are inferred, through a information system. This work suggests the development of flexible information systems and provides business managers with a tool to identify and implement new business strategies, based on user strategies

KEYWORDS

Information system, Business activity, User strategy, Knowledgebase, Supervised machine learning

1. INTRODUCCIÓN

Con el propósito de formalizar las intenciones de los usuarios de los sistemas de información empresarial, en el desarrollo de las actividades diarias de acuerdo con sus funciones y con el objetivo de mejorar su desempeño (Khodabandelou, Hug, y Salinesi, 2014), se propone un método para descubrir las estrategias que los usuarios utilizan para lograr sus objetivos personales, alineados con los objetivos del negocio. Este método (Figura 1) se basa en un algoritmo de aprendizaje automático supervisado (Khodabandelou, 2013), que se desarrolla en base a las actividades de los negocios en general y las reglas de un negocio específico. El algoritmo se entrena y aprende de una base del conocimiento (Flores *et al.*, 2010; Niu *et al.*, 2012; Pavón, 2016; De Sa *et al.*, 2017) o de un log de eventos de un negocio en general. Las reglas heurísticas de supervisión, se estructuran con base en las actividades prescritas y reglas del negocio. Para probar el algoritmo, se utiliza un conjunto de actividades prescritas del negocio específico (ventas). Finalmente, utilizando las palabras clave de un experto del negocio específico, se formalizan las intenciones del usuario.

La base del conocimiento contiene información empresarial, extraída de artículos de noticias y foros web. Se utiliza el conjunto de datos NewsIR'16 de Signal Media; Este conjunto de datos contiene un millón de artículos, sin embargo, para el presente trabajo, se selecciona una muestra de 200 artículos (suficientes para probar el método en un tiempo de procesamiento pertinente). La estructura de esta base del conocimiento consta de las siguientes tablas: artículos, oraciones, menciones_estrategias, estrategias_candidatos, características_estrategias, estrategias_usuario, estrategias_regla, estrategias_regla-0, pesos_estrategias, intenciones_usuario y log_actividades. Esta última tabla, se genera a partir de un log de eventos para un negocio en general y se obtuvo a través de un documento de la librería Kaggel.

A partir de los artículos se extraen las oraciones (frases separadas por un punto); se etiquetan sustantivos, adjetivos, verbos y adverbios (de acuerdo con el procesamiento del lenguaje natural NLP). Se extraen las menciones de los tokens (palabras etiquetadas) y se estructuran las oraciones según patrones de oración sujeto-verbo-objeto (en inglés) (Fromkin *et al.*, 2003). Estas oraciones estructuradas son las actividades del

negocio (estrategias del negocio). Finalmente se extraen las características semánticas de las estrategias del negocio en general; Con lo que se completa la estructura y contenidos de la base del conocimiento, por una parte; por otra parte y como fuente alternativa del conocimiento se cuenta con el log de eventos para un negocio en general.

Para el entrenamiento del algoritmo, se define el negocio específico (ventas), de acuerdo con el modelo de actividades que se presenta en la Figura 2, que sirve de base y de acuerdo con la experticia en el negocio de ventas, se establecen las reglas heurísticas de supervisión; en base de las cuales se generan las estrategias del usuario, categorizadas (estrategias_usuario: todas las estrategias extraídas, estrategias_regla: aquellas cuya probabilidad de supervisión es mayor a cero y estrategias_regla-0: aquellas que no se ajustan a las reglas heurísticas).

Se prueba el algoritmo con las estrategias del usuario (estrategias_regla) válidas para un negocio de ventas y con las actividades prescritas del negocio de ventas; en base de lo cual y de acuerdo con las palabras clave (venta, cliente, pedido, cotización, stock, precio, factura), se formalizan las intenciones del usuario de un negocio de ventas.

1.1. CONCEPTOS RELEVANTES

Evento. - Es un incidente que ocurre en el entorno empresarial, que está asociado a una determinada actividad y corresponde a una instancia de proceso (Khodabandelou *et al.* 2015).

Log de eventos. - Archivo plano que contiene los registros de eventos (actividades de proceso), registrados por sistemas de información, como producto del desarrollo de las actividades diarias del negocio (Khodabandelou 2013, Cheng y Kumar 2015; Khodabandelou *et al.*, 2014a; Khodabandelou *et al.*, 2013b).

Log de eventos genérico. - Registro de eventos que genera por defecto el sistema de información en cualquier negocio.

Log de eventos de calidad. - Se corrige y depura el log de eventos genérico, para un propósito específico y para mejorar la calidad de la minería de intenciones (Van Hee *et al.*, 2011; Suriadi *et al.*, 2017b; Van der Aalst, 2011).

Minería de intenciones. - Es el descubrimiento de estrategias de usuario (intenciones de usuario) a través de técnicas y herramientas de minería de procesos, aplicadas al log de eventos de calidad (Epure *et al.*, 2014; Khodabandelou *et al.*, 2015; Khodabandelou *et al.*, 2013a).

Intenciones del usuario. - Sin perder de vista las reglas del negocio y de acuerdo con la madurez y experiencia adquirida, el usuario realiza sus actividades diarias con la intención de alcanzar sus metas de acuerdo con los objetivos del negocio, tratando de ahorrar recursos y mejorar la calidad de los resultados. En la ejecución de una determinada instancia de proceso, el usuario desarrolla sus actividades, pero además, utiliza sus propias estrategias (intenciones), que le permiten cumplir con sus metas de manera más eficiente. La intención de cada usuario se puede cumplir con varias estrategias y cada estrategia se puede utilizar para cumplir con varias intenciones.

Procesos prescritos. - Estos son los procesos promulgados en el negocio para el cumplimiento de sus objetivos (Khodabandelou *et al.*, 2013b).

Procesos reales. - Estos son los procesos que realmente se ejecutan en el negocio (combinación de procesos prescritos y estrategias del usuario) (Khodabandelou *et al.*, 2013a).

Sistema de información. - Conjunto de procesos que se ejecutan de forma manual o computarizada en el desarrollo de las actividades empresariales diarias (Khodabandelou *et al.*, 2013a).

1.2. TRABAJOS PREVIOS

Técnicas y herramientas de minería de procesos. - Se utilizan como puente entre la minería de datos y el aprendizaje automático; en general, se utilizan para descubrir modelos de proceso orientados a las actividades del usuario, también para identificar estrategias de los usuarios. Además se utilizan

los modelos ocultos de Markov (HMM) en el descubrimiento de modelos de procesos intencionales (Khodabandelou *et al.*, 2014a).

Aprendizaje automático. - Este algoritmo se utiliza para clasificar secuencias de actividades en función de características similares. Sin embargo, también se podría utilizar un algoritmo de aprendizaje automático profundo para extraer las intenciones subyacentes en las huellas de los procesos (Khodabandelou *et al.*, 2014a).

Map Miner Method. - Este es el mejor producto desarrollado hasta el momento en este nuevo campo denominado minería de intención (Khodabandelou *et al.*, 2014a), permite construir modelos de procesos intencionales que podrían ser la base para futuras investigaciones en este campo; además, éste método permite definir y concretar estrategias y actividades de los usuarios en diferentes niveles de abstracción.

MAP. - Este es un lenguaje de modelado que se utiliza para obtener meta modelos de procesos intencionales que ayudan a formalizar modelos de procesos (Khodabandelou *et al.*, 2013b). No obstante, se podría usar lenguajes de modelado UML para hacer lo mismo, ya que es ampliamente conocido en el modelado de ingeniería de software.

Petri-Nets. - Esta herramienta se utiliza para representar modelos de procesos, en particular, modelos de procesos cíclicos (Khodabandelou *et al.*, 2013b). Para objetivo del presente trabajo, este recurso podría utilizarse para representar los modelos de estrategia de usuario.

Baum Welch. - Se utiliza este algoritmo para descubrir parámetros desconocidos en la aplicación de HMM (Khodabandelou *et al.*, 2013b). En el presente trabajo se podría utilizar para determinar los parámetros de mapeo de las estrategias e intenciones del usuario.

KAOS. - Esta herramienta permite modelar requisitos como instancias de un meta modelo conceptual que apoyan las intenciones de los usuarios, pero no se utiliza debido a su rígida definición de las tareas

(Khodabandelou *et al.*, 2013b). Sin embargo, esta herramienta ayudaría a definir niveles de abstracción en relación con las divisiones jerárquicas de las actividades y estrategias de los usuarios.

HMM. - Se utiliza para descubrir intenciones a partir de las trazas de procesos y para descubrir elementos como procesos, estrategias y modelos que están relacionados con la minería de intenciones y están presentes en el log de eventos.

ProM. - Esta es una herramienta de software libre, que se usa para hacer minería de procesos intencionales.

Khodabandelou (2013; 2014), propuso un método llamado minería de intención; este método está orientado a descubrir el modelo de proceso. Ella confía en el aspecto intencional de los procesos, que se basa en el mapa de meta modelo del proceso.

Epure (2014), desarrolló el Método Map Miner (MMM), este método genera modelos de procesos intencionales y en base a ellos, se emiten recomendaciones para desarrolladores de sistemas de información flexibles y conscientes de procesos.

Huang *et al.* (2018), en su trabajo, proponen la extracción de palabras y oraciones para formalizar las intenciones de los desarrolladores de software en línea (método similar al presente trabajo). No obstante, Huang desarrolla una técnica de minería intencional para emitir recomendaciones para los desarrolladores de software.

Cohan-Sujay y Madhulika (2012) proponen un método basado en la información que circula en las redes sociales, este método extrae oraciones simples y mediante machine learning, infiere la intención comercial (compra y venta) de los usuarios.

El presente trabajo se completa de la siguiente manera: la sección 2 enumera los materiales utilizados en el presente trabajo, la sección 3 describe el método basado en el aprendizaje automático supervisado, la

sección 4 presenta un análisis de los resultados, la sección 5 cubre la discusión y la sección 6 presenta la conclusión.

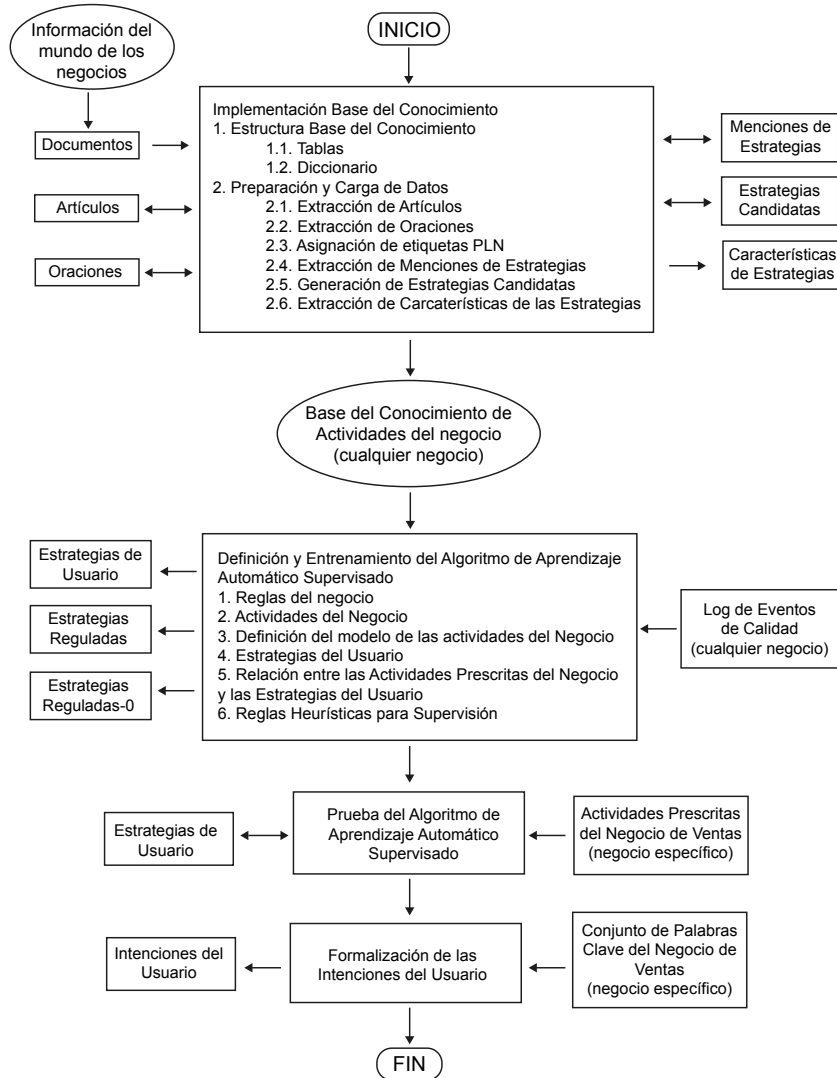


Figura 1. Flujoograma del Método.

Fuente: elaboración propia.

2. MATERIALES

Se utilizan varias herramientas de software, patrones y plantillas, datos y repositorios, y otros recursos, en la implementación del presente trabajo.

Sistema Operativo. - Ubuntu 16.04, especialmente porque es gratuito, seguro y exige pocos requisitos de sistema; además de su facilidad de uso.

Gestor de Base de Datos. - PostgreSQL 9.6, junto con pgAdmin III para almacenar los datos y administrar la base del conocimiento.

Lenguaje de Programación. - Python 3.6 para la codificación de los procedimientos y funciones de los algoritmos; además de la librería (plantilla) DeepDive y su requisito Docker para visualización.

Repositorios de Datos. - Foro web de noticias de Signal Media NewsIR'16 (documentos para alimentar la base del conocimiento) y kaggle dataset (log de eventos).

IEEE CIS. - Contexto BPM (Business Process Models) de descubrimiento de procesos y modelos de procesos para log de eventos.

Palabras Clave del Experto en Ventas. - Documentos de foros web (Wordstream, 2012; Sales & Marketing Keyword, 2020).

Programación del Lenguaje Natural (NLP: Natural Language Processing). - CoreNLP de la biblioteca digital de la Universidad de Stanford, con las siguientes funciones:

- Parsing, que en este trabajo se utiliza para definir las oraciones (frases separadas por un punto) contenidas en los artículos.
- Tokenizador, herramienta de software que extrae palabras o conjunto de palabras de un texto y las individualiza

- Lematizador, que unifica todas las formas de los verbos (modos y tiempos) a su forma en infinitivo.
- Etiquetador: Tokens, POS: Part Of Speech, NER: Named Entity Recognition.
- Conjunto de etiquetas de Penn TreeBank (Marcus *et al.*, 1993), utilizadas en la etiquetación de los tokens: sustantivos, adjetivos, verbos y adverbios, presentes en una oración.

3. MÉTODO

A continuación, se describe y explica punto por punto el método propuesto y que a manera de resumen se presenta en la Figura 1.

3.1. IMPLEMENTACIÓN DE LA BASE DEL CONOCIMIENTO

Una base del conocimiento, donde se almacena la información del negocio (Flores *et al.*, 2010), Sin importar el tipo de negocio y su dominio, que permite conocer, comprender y conceptualizar las actividades de un negocio, se crea la base del conocimiento, de acuerdo con las políticas y reglas del negocio (Ghasemi y Amyot, 2020); se especifican a continuación los pasos para crear una base de conocimientos para cualquier negocio en general.

3.1.1. ESTRUCTURA DE LA BASE DEL CONOCIMIENTO

Se estructura la base del conocimiento, de acuerdo con el método de diseño relacional de bases de datos (Díaz-Rodríguez, 2015). El escenario, es el mundo de los negocios multidisciplinarios. El recurso, las noticias que se publican en la web (Signal Media NewsIR'16). Las tablas y diccionario que se presentan a continuación, son el producto final de procesamiento y carga de datos que se realiza en los siguientes puntos del método propuesto en este trabajo y que se resume en el flujograma de la Figura 1.

- Tablas

articulos(doc_id, doc_text)

oraciones(doc_id, oracion_id, oracion_text, tokens, lemmas, pos_tags, ner_tags)

menciones_estrategias(doc_id, oracion_id, mencion_id, mencion_text, inicio_indice, fin_indice)

estrategias_candidatas(estrategia1_id, estrategia1_nombre, estrategia2_id, estrategia2_nombre)

caracteristicas_estrategias(estrategia1_id, estrategia2_id, caracteristica)

log_estrategias(estrategia1_nombre, estrategia2_nombre)

estrategias_usuario(estrategia1_id, estrategia2_id, usuario_id, etiqueta)

estrategias_regla(estrategia1_id, estrategia2_id, etiqueta, regla_id)

estrategias_regla-0(estrategia1_id, estrategia2_id, etiqueta, regla_id)

pesos_estrategias(estrategia1_nombre, estrategia2_nombre, peso)

intenciones_usuario(estrategia1_nombre, estrategia2_nombre, peso)

- Diccionario

articulos. - Tabla.

característica. - Una característica de la estrategia.

características_estrategias. – Tabla de las características de las estrategias.

doc_id. - Identificador del documento contenido en el artículo.

doc_text. - Contenido textual del documento.

estrategia1_id, estrategia2_id. - Identificador de una estrategia.

estrategias_candidatas. - Tabla de las posibles estrategias.

estrategias_regla. - Tabla de todas las estrategias con valor de probabilidad mayor a cero.

estrategias_regla-0. - Tabla de las estrategias excepto las que tienen valor de probabilidad igual a cero.

estrategias_usuario. - Tabla que contiene las estrategias del usuario.

fin_indice. - Posición donde termina la estrategia mencionada en la oración.

inicio_indice. - Posición donde comienza la mención de la estrategia en la oración.

lema. - Forma de una palabra que aparece como entrada en un diccionario y se utiliza para representar todas las demás formas posibles. Por ejemplo, el lema "jugar" representa "juegas", "juego", "jugó", etc.

log_estrategias. - Tabla, contiene las estrategias de usuario de un negocio específico.

men_id. - Identificador de la mención de una estrategia contenida en una oración.

mencion_texto. - Texto de la mención de una estrategia.

menciones_estrategias. - Tabla de actividades del negocio y estrategias de usuario extraídas de las oraciones.

ner_tags. - Lista de etiquetas de los nombres de entidades reconocidas.

nombre_estrategia1, nombre_estrategia2. - Nombre de una estrategia.

oracion_id. - Identificador de una oración.

oraciones. - Tabla de oraciones.

oraciones_texto. - Texto de una oración.

peso. - El peso asignado por la probabilidad de que una estrategia sea válida.

pesos_estrategias. - Tabla de las estrategias con su peso (valor de probabilidad).

pos_tags. - Lista de etiquetas de partes de texto.

regla_id. - Identificador de la regla heurística del negocio.

token. - Cada palabra o conjunto de palabras que compone un texto.

3.1.2. PREPARACIÓN Y CARGA DE DATOS

- Extracción de artículos

El texto fue extraído de 200 documentos que se publican en foros empresariales multidisciplinares. Y esa información se ha cargado en la tabla artículos.

articulos = extraer (<http://goo.gl/forms/5i4KldoWIX>)

- Extracción de oraciones

El contenido textual de cada documento es tokenizado (Hammond, 2020), es decir, el texto se divide en oraciones (parsing) y cada oración se divide en palabras (tokens). Las oraciones tokenizadas se almacenan en la tabla de oraciones.

oraciones = parsing (texto del documento), cada punto define a una oración

tokens = tokenizado (oración), cada palabra es un token

- Asignación de etiquetas NLP (Natural Language Processing)

A cada token de cada oración se le asigna una etiqueta como parte de un texto (POS: Part Of Speech), utilizando el sistema CoreNLP de Stanford (Manning *et al.*, 2015) y el estándar de etiquetas “The Penn TreeBank Tagset” (Marcus *et al.*, 1993).

token etiquetado = etiqueta POS (token)

Para identificar cada token (como un sustantivo, adjetivo, verbo o adverbio) y estructurar oraciones mínimas que representen las estrategias del usuario de los sistemas de información del negocio, se utiliza el sistema CoreNLP de Stanford en la asignación de etiquetas de Reconocimiento de Entidades Nombradas (NER: Named Entity Recognition) según la Tabla 1. Para esto, se utiliza el siguiente código Python.

```
pos[], ner[], k=0,
pos = extract_tokens(oraciones),
for i in pos:
    l = 0,
    for j in i:
        if j=='VB' or j=='VBD' or j=='VBG' or j=='VBN' or j=='VBP' or j=='VBZ':
            ner[k][l] = "ACTIVITY"
        else if j=='NN' or j=='NNS':
            ner[k][l] = "NOUN"
        else if j=='JJ':
            ner[k][l] = "ADJECTIVE"
        else if j=='JJR':
            ner[k][l] = "ADJECTIVE COMPARATIVE"
        else if j=='JJS':
            ner[k][l] = "ADJECTIVE SUPERLATIVE"
        else if j=='RB':
            ner[k][l] = "ADVERB"
        else if j=='RBR':
```

```

ner[k][l] ="ADVERB COMPARATIVE"
else if j=="RBS":
ner[k][l] ="ADVERB SUPERLATIVE"
l=l+1,
k=k+1

```

Tabla 1. Etiquetas de Tokens y Reconocimiento de Entidades Nombradas.

Etiqueta de Token	Nombre Entidades Reconocidas	Significado
VB, VBD, VBG, VBN, VBP, VBZ	ACTIVITY	Verbos en todos sus modos y tiempos
NN, NNS	NOUN	Sustantivos comunes
NNP, NNPS	PERSON / ORGANIZATION	Sustantivos
JJ	ADJECTIVE	Adjetivos
JJR	COMPARATIVE ADJECTIVE	Adjetivos comparativos
JJS	SUPERLATIVE ADJECTIVES	Adjetivos superlativos
RB	ADVERB	adverbios
RBR	ADVERB ADJECTIVE	Adverbios comparativos
RBS	ADVERB ADJECTIVES	Adverbios superlativos

Nota. La columna 1 contiene la etiqueta del token, la columna 2 contiene el nombre de la entidad reconocida en el texto y la columna 3 contiene el significado gramatical de la entidad.

Fuente: elaboración propia.

- Extracción de Menciones de Estrategias

De la tabla de oraciones, se extrajeron los tokens etiquetados que tienen secuencias correspondientes a las estructuras genéricas de oraciones que siguen la gramática inglesa (Fromkin *et al.*, 2003); de acuerdo con las siguientes opciones.

- “VERB”
- “NOUN” + “VERB”
- “NOUN” + “VERB” + “NOUN”

- “NOUN” + “VERB” + “ADJECTIVE”
- “NOUN” + “VERB” + “ADVERB”

Se genera la tabla menciones_estrategias con el siguiente código python.

```

num_tokens = len(ner_tags),
indices = (for i in num_tokens if (ner_tags[i] == “ACTIVITY”)
          and (i == 0 or ner_tags[i-1] != “ACTIVITY”)),
for inicio_indice in indices:
    fin_indice = inicio_indice,
    if inicio_indice > 0 and (ner_tags[inicio_indice-1] == “NOUN”
        or ner_tags[inicio_indice-1] == “PERSON”
        or ner_tags[inicio_indice-1] == “ORGANIZATION”):
        inicio_indice = inicio_indice - 1
    if fin_indice < num_tokens
        and (ner_tags[inicio_indice+1] == “ADJECTIVE”
            or ner_tags[inicio_indice+1] == “ADVERB”):
            fin_indice = fin_indice + 1
    mencion_id = (doc_id, sentence_index, inicio_indice, fin_indice),
    mencion_texto = join(map(strategy_mention i: tokens[i], xrange(inicio_indice,
        fin_indice + 1))),
    menciones_estrategias = (doc_id, sentence_id, mencion_id, mencion_texto, inicio_indice, fin_
indice)

```

Además, se depura la tabla menciones_estrategias eliminando elementos innecesarios para el propósito de este trabajo, por ejemplo, caracteres no imprimibles, elementos repetidos, sinónimos y otros

- Generación de posibles estrategias

Se genera la tabla `estrategias_candidatas` en base a las estrategias almacenadas en la tabla `menciones_estrategias`. Se toma cada par de estrategias y / o actividades relacionadas que se extrajeron de la misma oración, cuya distancia no es mayor a dos tokens y no se superponen. Para lo cual se utiliza el siguiente código python.

```
num_actividad(doc_id, oracion_indice, COUNT(p))=menciones_estrategias(p, _, doc_id, oracion_
indice,
                    →, →),
estrategias_candidatas(p1, p1_nombre, p2, p2_nombre) = num_actividad(doc_id, oracion_id, num_p),
                    menciones_estrategias(p1, p1_nombre, doc_id, oracion_id, p1_inicio, →),
                    menciones_estrategias(p2, p2_nombre, doc_id, oracion_id, p2_inicio, →),
                    num_p < 3,
                    p1_nombre != p2_name,
                    p1_inicio != p2_inicio
```

- Extracción de Características de las Estrategias

Finalmente, se extraen las características de las posibles estrategias válidas y se almacenan en la tabla `caracteristicas_estrategias`; para esto se utiliza el siguiente código python.

```
significados = [], tokens= extract_token(oraciones)
for i, t in tokens:
    sent.append(palabra=t, lemma=lemmas[i], pos=pos_tags[i], ner=ner_tags[i]),
p1_pnt = (p1_inicio_indice, length=(p1_fin_indice-1_inicio_indice+1)),
p2_pnt = (p2_inicio_indice, length=(p2_fin_indice-2_inicio_indice+1)),
for caracteristica in (significados, p1_pnt, p2_pnt):
    caracteristicas_estrategias=yield [p1_id, p2_id, caracteristica]
```

3.2. DEFINICIÓN DEL ALGORITMO DE APRENDIZAJE AUTOMÁTICO SUPERVISADO

La base del conocimiento creada en los pasos previos, contiene los datos y la información de cualquier negocio. Para el caso específico, se considera un negocio de ventas; por lo tanto, el algoritmo de aprendizaje automático supervisado está diseñado con los pasos que se especifican a continuación.

3.2.1. REGLAS DEL NEGOCIO

La base de conocimiento que está estructurada y alimentada con datos de carácter general contiene las actividades para satisfacer las reglas de cualquier negocio, se pueden ver ejemplos de reglas para cualquier negocio en (Rowland, 2004). No obstante, para el caso del negocio de ventas, se definen las siguientes reglas.

- Para ordenar o cotizar, el cliente debe estar registrado
- Un pedido o cotización, se puede completar con stock local o remoto
- Un pedido o cotización no confirmados, compromete el stock por un tiempo determinado, de acuerdo con las políticas del negocio.
- Un pedido o cotización confirmada que se registra como una venta afecta el stock local y/o remoto
- Se puede enviar una venta a domicilio
- Cada vendedor (usuario) puede usar sus propias estrategias para mejorar su desempeño, mientras observa las políticas y reglas comerciales

3.2.2. ACTIVIDADES DEL NEGOCIO

Se especifican las actividades prescritas del negocio de ventas (Pérez, Yzquierdo, y Silverio, 2015). El modelo de actividades de este negocio se puede ver en la Figura 2; donde se pueden validar las mencionadas actividades prescritas (procesos prescritos) del negocio de ventas en cuestión. Este modelo

de la Figura 2, ha sido elaborado con el listado de actividades que muestra a continuación y utilizando el algoritmo de minería de procesos “Alpha Miner” disponible en el conjunto de la herramienta ProM 6.9.

- Atención al Cliente
 - Crear cliente
 - Activar cliente
 - Actualizar cliente
- Generar pedido de cliente
 - Pedido completo de cliente
 - Cotización de pedido de cliente
- Emitir cotización de cliente
 - Control local de existencias
 - Control remoto de stock
 - Registrar artículos
- Pedido completo de cliente
 - Cotización de cliente emitida
 - Pedido del cliente confirmado
 - Envío de pedido de cliente
- Control local de existencias
 - Verificar stock
 - Emitir alerta de stock mínimo local

- Stock comprometido
- Control remoto de stock
 - Verificar stock
 - Emitir alerta de stock mínimo remoto
 - Stock comprometido
- Envío de pedido de cliente
 - Entrega de pedidos con stock local
 - Entrega de pedidos con stock remoto
 - Entrega local del pedido del cliente
 - Emitir orden de entrega remota
 - Entrega a domicilio del pedido
- Cancelar pedido de cliente
 - Liberar existencias comprometidas
- Registro de venta
 - Verificar orden confirmada
 - Registrar descuentos
 - Verificar impuestos
 - Registrar método de pago
 - Facturación
- Enviar venta a domicilio

- Verificar la dirección del cliente
- Seguimiento de entrega a domicilio
- Confirmar entrega a domicilio
- Récord de ventas
 - Por producto (descuento por: liquidación, lanzamiento)
 - Por cliente (descuento para: cliente frecuente)
 - Por almacén
 - Por período / temporada
 - Por método de pago (descuento actual)
 - Por promoción
 - Por vendedor
 - Por método de entrega
 - Por volumen de ventas (descuento por importe de venta)
 - Entre otras dimensiones

3.2.3. DEFINICIÓN DEL MODELO DE ACTIVIDADES DEL NEGOCIO

Se genera el modelo de actividad del negocio (Tece, 2010; Foss y Saebi, 2018), de acuerdo con las actividades prescritas del negocio de ventas, y utilizando el algoritmo Alpha Miner de ProM. Este modelo se muestra en la Figura 2.

3.2.4. ESTRATEGIAS DEL USUARIO

Las actividades que ejecuta el usuario en el desarrollo de sus tareas cotidianas en el negocio, no todas están de acuerdo con los procesos prescritos del negocio (Teece, 2010). El usuario con el propósito de mejorar su desempeño, utiliza sus propias actividades a las que se denomina estrategias del usuario. En el presente trabajo, estas estrategias de usuario son extraídas de la base del conocimiento, estructurada en los pasos previos del presente método. Las mencionadas estrategias de usuario están almacenadas en la tabla `estrategias_usuario`.

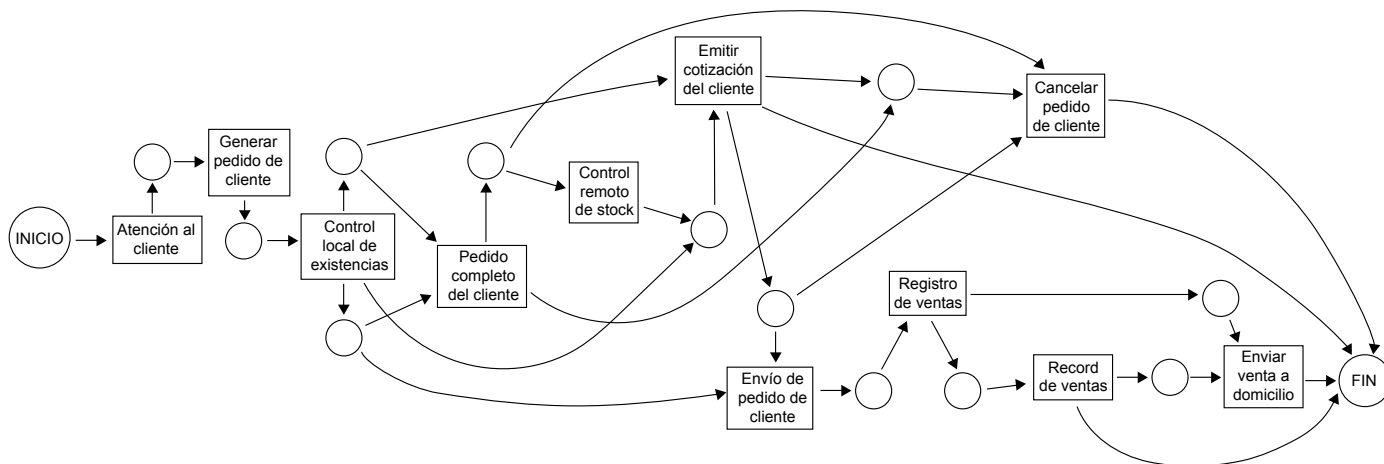


Figura 2. Modelo de actividades del negocio de ventas.

Fuente: elaboración propia.

3.2.5. RELACIÓN ENTRE LAS ACTIVIDADES PRESCRITAS DEL NEGOCIO Y LAS ESTRATEGIAS DEL USUARIO

Se presenta a continuación, el análisis de las actividades y su mapeo con las estrategias del usuario, según el modelo de negocio de ventas (Figura 2).

- Una actividad del negocio, por sí sola, puede ser una estrategia de usuario.

- Una actividad del negocio puede formar parte de una o varias estrategias de usuario.
- Una estrategia de usuario puede estar compuesta por una o más actividades del negocio.
- Una estrategia de usuario puede estar compuesta por actividades prescritas del negocio y/o no prescritas, pero del mismo dominio del negocio.

3.2.6. REGLAS HEURÍSTICAS PARA SUPERVISIÓN

De la heurística aplicada en los pasos previos del presente método, en la estructuración de la base del conocimiento para un negocio cualquiera en general; y, con base en el modelo de las actividades prescritas del negocio de ventas (Figura 2). Se deducen las siguientes reglas heurísticas para el negocio específico de ventas (Makings y Barnard, 2014).

- Las actividades que son candidatas a estrategias, etiquetadas como “ACTIVITY” y que están demasiado separadas (más de 2 tokens en el medio) en una oración, se marcan como falsas (negativas).
- Las actividades de las candidatas para estrategias etiquetadas como “ACTIVITY” precedidas por un sustantivo, etiquetadas como “NOUN” o “PERSON” u “ORGANIZATION” en una oración, correspondientes al patrón de sujeto + verbo, se marcan como verdaderas (positivas).
- Las actividades que son candidatas para estrategias con el patrón de oración sujeto + verbo, y que están precedidas por un adjetivo etiquetado como “ADJECTIVE” en una oración, correspondiente al patrón sujeto + verbo + adjetivo, se marcan como verdaderas (positivas).
- Las actividades que son candidatas para estrategias con el patrón de oración sujeto + verbo, y que están precedidas por un adverbio etiquetado como “ADVERB” en una oración, correspondiente al patrón sujeto + verbo + adverbio, se marcan como verdaderas (positivas).

- Las actividades candidatas para estrategias que contienen palabras relacionadas con el dominio del negocio (dealing: “order”, “quotation”, “stock”, “sale”, “price”, “customer”, “user”) se marcan como verdaderas.

El código Python que se muestra a continuación, permite calcular el peso de cada par de estrategias con respecto al cumplimiento de las reglas heurísticas.

```

TRADING = frozenset(["sell", "buy"]),
DEALING = frozenset(["order", "quotation", "stock", "sale", "price", "customer", "user"]),
MAX_DIST = 2,
p1_start_idx = min(p1_begin, p2_begin),
p1_end_idx = min(p1_end, p2_end),
p2_start_idx = max(p1_begin, p2_begin),
p2_end_idx = max(p1_end, p2_end),
p1_lemmas = lemmas[p1_start_idx:p1_end_idx],
intermediate_lemmas = lemmas[p1_end_idx+1:p2_start_idx],
p2_ner_tags = ner_tags[p2_start_idx:p2_end_idx],
estrategias_regla-0(estrategia1, estrategia2,_),
estrategias_regla(estrategia1, estrategia2,_),
if len(intermediate_lemmas) > MAX_DIST:
    estrategias_regla-0=yield [estrategia1, estrategia2, 'negative:far_apart']
if p1_end_idx > p1_start_idx and 'NOUN' in p2_ner_tags:
    estrategias_regla=yield [estrategia1, estrategia2, 'positive:noun-verb-noun']
if p1_end_idx > p1_start_idx and 'ADJECTIVE' in p2_ner_tags:
    estrategias_regla=yield [estrategia1, estrategia2, 'positive:noun-verb-adjective']

```

if p1_end_idx > p1_start_idx and 'ADVERB' in p2_ner_tags:

estrategias_regla=yield [estrategia1, estrategia2, 'positive:noun-verb-adverb']

if len(TRADING.intersection(p1_lemmas)) > 0:

estrategias_regla=yield [estrategia1, estrategia2, 'positive:sell_buy_activity']

if len(DEALING.intersection(p1_lemmas)) > 0:

estrategias_regla=yield [estrategia1, estrategia2, 'positive:dealing_activities']

Tabla 2. Estrategias de usuario ponderadas.

ESTRATEGIA 1	ESTRATEGIA 2	PESO
customer quotation	stock control	0.946
price sale	stealthy	0.832
price sale	trusted	0.832
price sale	seductive	0.811
selloff was	stocks pulled	0.777
selloff was	cut	0.777
opinion tells	stocks are	0.772
stock has	moving average	0.736
stock has	day moving	0.736
best-selling	billon	0.716
best-selling	grow	0.716
order makes	be	0.715
order makes	hearing has	0.715
Be	price matched	0.685
by customer	had	0.624
by customer	sales record	0.624
have	prices change	0.575
price does	tell	0.571
Cordero said	people respond	0.499
Cordero said	hopes	0.499
best-selling	bonanza	0.444
Bunning has	order saying	0.175

judge has	ordered court-appointed	0.018
-----------	-------------------------	-------

Nota. La columna de ponderación 3 representa la probabilidad de que las estrategias (columna 1 y 2), juntas, puedan inferir la intención del usuario.

Fuente: elaboración propia.

En la Tabla 2, se presentan los pares de estrategias obtenidas de la tabla estrategias_regla ordenadas por el peso, de acuerdo con el cumplimiento de las reglas heurísticas.

3.3. DEFINICIÓN DEL CONJUNTO DE DATOS DE ENTRENAMIENTO

Las estrategias del usuario que son en sí, el conjunto de datos de entrenamiento del algoritmo, se obtienen por una parte; de la tabla estrategias_usuario, generada en el proceso de estructuración de la base del conocimiento. Por otra parte y alternativamente, las estrategias del usuario se pueden obtener del log de eventos extraído del repositorio kaggle. Para el caso que se desarrolla en el presente trabajo, se utilizó las estrategias del usuario contenidas en la base del conocimiento; no obstante, para el caso del log de eventos, se resume a continuación el proceso que se debe seguir (Andrews *et al.*, 2018).

- Obtener el log de eventos genérico (LEG).
- Generar un log de eventos de calidad (LEC) a partir del LEG.
- Del LEC, extraer las oraciones y aplicar los pasos del subtítulo “3.1.2. PREPARACION Y CARGA DE DATOS”, para generar las estrategias del usuario.

3.4. DEFINICIÓN DEL CONJUNTO DE PALABRAS CLAVE DEL NEGOCIO (DATOS DE PRUEBA)

El valor de la variable de predicción (estrategia de usuario), se lo define en función de las actividades contenidas en el LEC y se lo obtiene como se describe a continuación.

- Una actividad, como el patrón de la oración, “verbo”, en este caso, el sujeto y el objeto son tácitos.
- Un sustantivo y un verbo, como patrón de oración, “sustantivo + verbo”

- Un sustantivo y un verbo, y otro sustantivo, como patrón de oración, “sustantivo + verbo + sustantivo”
- Un sustantivo, un verbo y un adjetivo, como patrón de oración, “sustantivo + verbo + adjetivo”
- Un sustantivo, un verbo y un adverbio, como patrón de oración, “sustantivo + verbo + adverbio”

De acuerdo con los patrones anteriores (Fromkin *et al.*, 2003), cada palabra clave podría ser un sustantivo, un verbo, un adjetivo o un adverbio. Para este trabajo en el que se abordó el negocio de ventas, las palabras clave más importantes son: “order”, “quotation”, “stock”, “sale”, “price”, “customer”, “user”.

3.5. FORMALIZACIÓN DE LAS INTENCIONES DEL USUARIO

Finalmente, con base en las estrategias almacenadas en la tabla estrategias_usuario y las estrategias emparejadas y cuyo peso es mayor que cero (estrategias_regla), como se muestran en la Tabla 2; además del criterio del experto en el negocio de ventas (MYOB Enterprice Solution, 2020; Vasilev y Kehayova-Stoycheva, 2017), se infieren las intenciones del usuario como se muestran en la Tabla 3.

Tabla 3. Intenciones del usuario.

ESTRATEGIA 1 USUARIO	ESTRATEGIA 2 USUARIO	INTENCIÓN DEL USUARIO
customer quotation	stock control	A quotation based on stock control
price	stealthy	Always is opportune, a good price
	trusted	
	seductive	
opinion tells	stocks are	The stock control is a factor critic of successful
best-selling	billon	A best-selling, always is healthy
	grow	
	bonanza	
be	price matched	The price must be the same to the competition
by customer	sales record	Know very well to the client
have	prices change	Maintaining variable prices is not an advantage

Nota. La columna 1 contiene una estrategia de usuario; la columna 2 contiene otra estrategia de usuario y la columna 3 contiene la intención del usuario inferida por las estrategias de usuario.

Fuente: elaboración propia.

4. ANÁLISIS DE RESULTADOS

De los 200 artículos que se procesaron, se obtuvo 5.149 oraciones. De estas oraciones; se extrajeron 16.943 menciones de actividades de negocios en general. Se clasificaron estas menciones de acuerdo con las reglas de un negocio de ventas y se obtuvieron 8.455 menciones de estrategias candidatas. A partir de estas menciones de estrategias candidatas para un negocio de ventas se generaron 7.930.978 características de las estrategias. Por otro lado, a partir del repositorio kaggle se generó el log de eventos de calidad del que se extrajeron 482 estrategias de usuario. A continuación se muestran los resultados del proceso de aprendizaje y supervisión del algoritmo, aplicado a la base del conocimiento de los negocios en general y las reglas de un negocio específico (ventas) en particular (Makings y Barnard, 2014).

4.1. SUPERVISIÓN A TRAVÉS DE REGLAS HEURÍSTICAS

Las estrategias ponderadas de usuario (Tabla 2) que se encuentran almacenadas en la tabla *estrategias_regla*; y cuyo peso se calcula, de acuerdo con el grado de emparejamiento, que el algoritmo de aprendizaje ha logrado establecer, entre cada par de estrategias del usuario; en base al porcentaje de cumplimiento de las reglas heurísticas, la validación de las estrategias del usuario de un negocio de ventas en particular, con respecto a las estrategias de usuario contenidas en la base del conocimiento de un negocio en general. A continuación se muestra un resumen del cumplimiento de las reglas heurísticas.

- Resultado positivo de reglas heurísticas
 - ***positivo: sale_buy_activity.*** - 1 Estrategia que se corresponde con las actividades de venta y compra
 - ***positivo: dealing_activities.*** - 35 Estrategias que se corresponden con las actividades de negociación.
 - ***positivo: noun-verb-noun.*** - 538 Estrategias que se corresponden con el patrón sustantivo-verbo-sustantivo de una oración

- **positivo: noun.** - 0 Estrategias
- **positivo: noun-verb.** - 0 Estrategias
- **positivo: noun-verb-adjective.** - 0 Estrategias
- **positivo: noun-verb-adverb.** - 0 Estrategias
- Resultado negativo de reglas heurísticas
 - **negativo: far_apart.** - 6706 Estrategias, cuyos tokens en la oración están separadas entre sí, una distancia superior a 2 tokens

4.2. CLASIFICACIÓN DE LAS ESTRATEGIAS DEL USUARIO

La tabla `estrategias_usuario` contiene todas las estrategias extraídas de la base de conocimiento, se validan contra las actividades y estrategias del negocio de ventas y las estrategias validadas se almacenan en la tabla `log_strategias`. Cada fila en la tabla `estrategias_usuario`, tiene una etiqueta (verdadero, falso o sin etiqueta), de acuerdo con el cumplimiento de las reglas heurísticas de supervisión. El “sin etiqueta”, significa que la estrategia del usuario no tiene correspondencia con el dominio del negocio de ventas. La etiqueta “falso”, significa que, individualmente, cada estrategia está relacionada con el negocio de ventas; pero juntas (`estrategia1` y `estrategia2`), no tienen relación con el dominio del negocio de ventas. La etiqueta “verdadero” significa que, tanto individualmente como en conjunto, las estrategias de usuario pertenecen al dominio del negocio de ventas. En consecuencia, las estrategias de usuario con etiqueta “verdadero” se almacena en la tabla `estrategias_regla`, que contiene las estrategias emparejadas y su respectivo peso mayor que cero.

4.3. INFERENCIA DE LAS INTENCIONES DEL USAURIO

De acuerdo con la Tabla 3, y desde el punto de vista de un experto en negocios de ventas (MYOB Enterprise Solution, 2020; Vasilev y Kehayova-Stoycheva, 2017), se analizan las intenciones del usuario, observando las recomendaciones de (Tan *et al.*, 2015).

- ***A quotation based on stock control.*** - En un negocio de ventas, el control de stock siempre será muy importante. Para una cotización, es vital definir tiempos para comprometer y liberar el stock (local o remoto).
- ***Always is opportune, a good price.*** - En este caso, el conocimiento de las estaciones y sus características en el sitio geográfico del negocio es determinante.
- ***The stock control is a factor critic of successful.*** - La implementación de alertas, para stocks mínimos y máximos, es fundamental en el mantenimiento de este factor crítico de éxito.
- ***Determinar un producto más vendido siempre es saludable.*** - Para cualquier negocio de ventas, el conocimiento de los productos “estrella”, permite una mejor administración de la fuerza de ventas y el consecuente beneficio para el negocio.
- ***The price must be the same to the competition.*** - Mantener precios acordes a los competidores es la estrategia más aconsejable, para no romper el mercado y mantener el equilibrio.
- ***Know very well to the client.*** - No hay nada mejor para el negocio de ventas que conocer al comprador
- ***Maintaining variable prices is not an advantage.*** - Definitivamente, no mantener un estándar de precio, no es saludable para el negocio.

5. DISCUSIÓN

5.1. RECURSOS

Los datos utilizados en este trabajo se adquirieron de foros web gratuitos (Signal Media, BigQuery Google, kaggle, The Penn Treebank, MYOB, Wordstream, entre otros), el repositorio académico Stanford NPL Group y datos tabulados del autor Wil van der Aalst. Se estructuró una base del conocimiento a partir

de archivos planos, observando los modelos de bases de datos relacionales y dimensionales. No se han utilizado datos de Organizaciones ni públicas ni privadas.

5.2. BASE DEL CONOCIMIENTO

La base del conocimiento que se crea en este trabajo está relacionada con el dominio empresarial, es decir, para cualquier negocio en general. Este conocimiento es generado y aprendido por el algoritmo de aprendizaje automático supervisado (Vasilev y Kehayova-Stoycheva, 2017). Como se puede apreciar en el método, a través del cual, se crea la base de conocimiento, los datos que se cargan, provienen de artículos publicados en noticias del campo de los negocios multidisciplinares.

5.3. SUPERVISED MACHINE LEARNING ALGORITHM

El algoritmo desarrollado, construye su aprendizaje en base a la base de conocimiento (actividades para un negocio en general). El algoritmo es entrenado para un negocio de ventas en particular; con base en las reglas del negocio de ventas (negocio específico) y las reglas heurísticas de supervisión. Para probar el algoritmo, se utiliza un conjunto de palabras clave (actividades y estrategias) del negocio de ventas. Como resultado de la prueba del algoritmo se obtienen las estrategias ponderadas del usuario; en base de las cuales se infieren las intenciones del usuario.

5.4. ESTRATEGIAS DEL USUARIO

Para cualquier negocio en general; a partir de las actividades almacenadas en la base del conocimiento, se obtienen las estrategias del usuario para un negocio en general. Con base en el modelo de actividades del negocio de ventas y las reglas heurísticas de supervisión, se generan las estrategias del usuario de ventas y se ponderan en base al cumplimiento de las condiciones de las reglas heurísticas.

5.5. USER INTENTIONS

Sin perder de vista las reglas del negocio y sin afectar las actividades prescritas (Vasilev y Kehayova-Stoycheva, 2017), los usuarios realizan sus actividades diarias, tratando de alcanzar un nivel óptimo de desempeño; pero a menudo, los sistemas de información no tienen la suficiente flexibilidad para lograr el objetivo de esas actividades. Entonces, los usuarios utilizan sus propias estrategias, con la intención de lograr sus objetivos personales, siempre alineados con los objetivos del negocio.

En el presente trabajo, en base a las estrategias del usuario, se infieren las intenciones del usuario. Un experto en el negocio puede interpretar esas intenciones del usuario y puede ayudar a identificar el comportamiento del usuario.

6. CONCLUSIONES

En todo proceso de minería, el objetivo es obtener un patrón (la pepita de oro) que represente el universo del recurso minado (datos, procesos, gráficos, mensajes, textos, documentos, etc). En este caso, se han extraído las intenciones del usuario, respecto al desarrollo de sus actividades en el negocio a través de un sistema de información. Esto podría usarse para determinar el comportamiento del usuario y mejorar su desempeño; con la ventaja que el método propuesto, presenta la alternativa de trabajar a partir de archivos planos o estructurados; el método permite además, estructurar una base del conocimiento que se podría generalizar para cualquier tipo de negocio.

REFERENCIAS BIBLIOGRÁFICAS

Andrews, R., Suriadi, S., Ouyang, C., y Poppe, E. (2018). Towards Event Log Querying for Data Quality: Let's Start with Detecting Log Imperfections. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-02610-3_7

- Cohan-Sujay, C., y Madhulika, Y.** (2012). Intention Analysis for Sales, Marketing and Customer Service. *COLING Demos* (December 2012), 33–40. <http://acl.eldoc.ub.rug.nl/mirror/C/C12/C12-3005.pdf>
- De Sa, C., Ratner, A., Ré, C., Shin, J., Wang, F., Wu, S., y Zhang, C.** (2017). Incremental knowledge base construction using DeepDive. *VLDB Journal*, 26(1), 81–105. <https://doi.org/10.1007/s00778-016-0437-2>
- Díaz-Rodríguez, O.** (2015). Methodology for Designing Relational Databases Based on Scenario Analysis Their Policies and Business Rules. *Tic*, 344(9), 77–88. <https://doi.org/10.17993/3c-tic.2015.42.197-209>
- Epure, E. V., Hug, C., Deneckère, R., y Brinkkemper, S.** (2014). What shall I do next? Intention mining for flexible process enactment. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8484, 473–487. https://doi.org/10.1007/978-3-319-07881-6_32
- Flores, I., Núñez, H., Ramos, E., Vitelli, J., y Rodríguez, V.** (2010). Construcción de bases de conocimiento en el dominio de la microbiología mediante técnicas de aprendizaje automático. *Conference: III Congreso Español de Informática. Actas V Simposio de Teoría y Aplicaciones de Minería de Datos, TAMIDA'2010*, Valencia, España. CEDI2010_ Bacilos (September 201), 315-323. <https://www.researchgate.net/publication/267751909>
- Foss, N. J., y Saebi, T.** (2018). Business models and business model innovation: Between wicked and paradigmatic problems. *Long Range Planning*, 51(1), 9–21. <https://doi.org/10.1016/j.lrp.2017.07.006>
- Fromkin, V., Rodman, R., y Hyams, N.** (2003). An Introduction to Language seventh edition. *The Sentence Patterns of Language* (pp. 117-165). <http://www.wjh.harvard.edu/~pal/pdfs/pdfs/7th.pdf>

- Ghasemi, M., y Amyot, D.** (2020). From event logs to goals: a systematic literature review of goal-oriented process mining. *Requirements Engineering*, 25(1), 67–93. <https://doi.org/10.1007/s00766-018-00308-3>
- Hammond, M.** (2020). *NLTK. Python for Linguists*, 291–296. <https://doi.org/10.1017/9781108642408.013>
- Huang, Q., Xia, X., Lo, D., y Murphy, G. C.** (2018). Automating Intention Mining. *IEEE Transactions on Software Engineering*, PP(c), 1-22. <https://doi.org/10.1109/TSE.2018.2876340>
- Khodabandelou, G.** (2013). Contextual recommendations using intention mining on process traces: Doctoral consortium paper. *Proceedings - International Conference on Research Challenges in Information Science*, pp. 115-122. <https://doi.org/10.1109/RCIS.2013.6577728>
- Khodabandelou, G., Hug, C., Deneckere, R., y Salinesi, C.** (2013). Supervised intentional process models discovery using Hidden Markov models. *Proceedings - International Conference on Research Challenges in Information Science*, pp. 225-2236. <https://doi.org/10.1109/RCIS.2013.6577711>
- Khodabandelou, G., Hug, C., Deneckère, R., y Salinesi, C.** (2014). Unsupervised discovery of intentional process models from event logs. *11th Working Conference on Mining Software Repositories, MSR 2014 - Proceedings*, pp. 282–291. <https://doi.org/10.1145/2597073.2597101>
- Khodabandelou, G., Hug, C., y Salinesi, C.** (2014). A novel approach to process mining: Intentional process models discovery. *Proceedings - International Conference on Research Challenges in Information Science*, pp. 78-89. <https://doi.org/10.1109/RCIS.2014.6861040>
- Makings, A., y Barnard, B.** (2014). The heuristics of entrepreneurs. 1. Introduction. 2. Literature Review. *Research Gate*, 1–37. <https://doi.org/10.2139/ssrn.3342170>

- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., y McClosky, D.** (2015). Stanford CoreNLP bindings for other languages. *The Stanford CoreNLP Natural Language Processing Toolkit*. <https://doi.org/10.3115/v1/p14-5010>
- Marcus, M., Santorini, B., Ann Marcinkiewicz, M., y Large, B.** (1993). *Building a Large Annotated Corpus of English: The Penn Treebank Building a Large Annotated Corpus of English*. The Penn Treebank Recommended Citation Recommended Citation. https://repository.upenn.edu/cis_reports/237
- MYOB Enterprise Solutions.** (2020). *MYOB EXO Business White Paper. M470 – Sales Analysis*. https://www.uil.co.nz/help/MYOBEXOBUSINESS_SalesAnalysisMatrix_whitepaper.pdf
- Niu, F., Zhang, C., Ré, C., y Shavlik, J.** (2012). Elementary: Large-scale knowledge-base construction via machine learning and statistical inference. *International Journal on Semantic Web and Information Systems*, 8(3), 42–73. <https://doi.org/10.4018/jswis.2012070103>
- Pavón, F.** (2016). *Generación de Conocimiento basado en Aprendizaje Automático y Aplicación en Diferentes Sectores*. UNED. Universidad Nacional de Educación a Distancia (España). <https://dialnet.unirioja.es/servlet/tesis?codigo=66958>
- Pérez, D., Yzquierdo, R., y Silverio, R.** (2015). Utilización de técnicas de minería de proceso en el entorno empresarial cubano. *Research Gate*, (August 2017), 1–15. <https://www.researchgate.net/publication/237843306>
- Rowland, E.** (2004). Principles of the business rule approach. *International Journal of Information Management*, 24(2), 196–197. <https://doi.org/10.1016/j.ijinfomgt.2003.12.007>
- Sales & Marketing Keyword.** (2020). *Columbia Greene Works. List of Resume Keywords and Competencies for Sales and Marketing Available*. https://www.resumeworld.ca/resume_best_practices/resume-keywords/sales-resume-keywords.html

- Tan, Y., Shi, Y., Buarque, F., Gelbukh, A., Das, S., y Engelbrecht, A.** (2015). Advances in Swarm and Computational Intelligence. En *6th International Conference, ICSI 2015 held in conjunction with the Second BRICS Congress, CCI 2015 Beijing, China, June 25-28, 2015 Proceedings, Part III. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9142(July 2016). <https://doi.org/10.1007/978-3-319-20469-7>
- Teece, D. J.** (2010). Business models, business strategy and innovation. *Long Range Planning*, 43(2–3), 172–194. <https://doi.org/10.1016/j.lrp.2009.07.003>
- Vasilev, J., y Kehayova-Stoycheva, M.** (2017). Sales analysis by the rectangle method. *Leonardo Electronic Journal of Practices and Technologies*, (30), 149–160. <http://lejpt.academicdirect.org>
- Wordstream.** (2012). *Sales Keyword*. <https://www.wordstream.com/popular-keywords/sales-keywords>