A successful case of collaboration between a company and research institutes for intelligent recommendation of items on an online retail portal.

Sebastián Vallejos[1], Luis Berdun[1], Marcelo Armentano[1], Silvia Schiaffino[1]
Sandra González Císaro[2], Oscar Nigro[2], Ignacio Cuesta[3] y Leonardo Balduzzi[3]

1 ISISTAN (CONICET - UNCPBA) - Campus Universitario, Paraje Arroyo Seco, Tandil, Argentina
{sebastian.vallejos, luis.berdun, marcelo.armentano, silvia.Schiaffino}@isistan.unicen.edu.ar
2 INTIA, Fac. Cs. Exactas - Campus Universitario, Paraje Arroyo Seco, Tandil, Argentina
{sagonci,onigro}@exa.unicen.edu.ar
3 Región Global, Tandil
{ignacio.cuesta, leonardo.balduzzi}@regionglobal.com

Resumen. This article describes the joint experience carried out to develop a prototype of an intelligent assistant aiming at bringing together sellers and buyers in an online sales portal. Originally, the project sought to carry out an intelligent characterization of users with their short-term preferences and their geospatial features. The project aimed to provide intelligent alerts about items potentially interesting to users, always considering that it is crucial that the information is acquired in a non-intrusive and intelligent way. With the synergy achieved between the parties from the joint development of activities, many windows of opportunity arose with the knowledge of the portal analysis by the researchers and with the knowledge of Artificial Intelligence techniques by the company. During the joint work, the different objectives were adapted in order to cover the effective needs of the company, achieving successful collaboration between the parties.

Un caso de éxito en la colaboración entre empresa e institutos de investigación para la recomendación inteligente de anuncios en un portal de venta online.

Sebastián Vallejos[1], Luis Berdun[1], Marcelo Armentano[1], Silvia Schiaffino[1]
Sandra González Císaro[2], Oscar Nigro[2], Ignacio Cuesta[3] y Leonardo Balduzzi[3]

1 ISISTAN (CONICET - UNCPBA) - Campus Universitario, Paraje Arroyo Seco, Tandil, Argentina
{sebastian.vallejos, luis.berdun, marcelo.armentano, silvia.Schiaffino}@isistan.unicen.edu.ar
2 INTIA, Fac. Cs. Exactas - Campus Universitario, Paraje Arroyo Seco, Tandil, Argentina
{sagonci,onigro}@exa.unicen.edu.ar
3 Región Global, Tandil
{ignacio.cuesta, leonardo.balduzzi}@regionglobal.com

Resumen. En el siguiente artículo se describe la experiencia conjunta realizada para el desarrollo de un prototipo de asistente inteligente que acercará a vendedores y compradores en un portal de ventas online. Originalmente el proyecto buscaba realizar una caracterización inteligente del usuario con sus preferencias a corto plazo y las características geoespaciales que lo definan. El proyecto buscó brindar alertas inteligentes sobre artículos de potencial interés para los usuarios, siempre considerando que es crucial que la información sea adquirida de una forma no intrusiva e inteligente. Con la sinergia alcanzada entre las partes a partir del desarrollo conjunto de actividades, se fueron apreciando muchas ventanas de oportunidades que surgieron con el conocimiento del análisis del portal por parte de los investigadores, y con el conocimiento de las técnicas de Inteligencia Artificial por parte de la empresa. Durante el trabajo realizado en conjunto se fueron adaptando los distintos objetivos a fin de poder abarcar las necesidades efectivas de la empresa, logrando una colaboración exitosa entre las partes.

1 Introducción

El Comercio electrónico ha evolucionado hasta convertirse en la actualidad en una de las principales opciones de venta. Según la Cámara Argentina de Comercio Electrónico (CACE) en el año 2021 la facturación del comercio electrónico creció en un 68% respecto al año 2020 logrando una facturación total de $1.520.000 millones de pesos. Para tener un impacto real del crecimiento sostenido en el tiempo, en el año 2018 según la CACE la facturación fue de $229.760 millones de pesos, es decir que en 3 años se multiplicó por casi 7 veces la facturación. Parte de este crecimiento se explica porque en el año 2020, durante la pandemia por Covid-19 se produjo un boom de compras online. Sin embargo, según el estudio realizado por la CACE en el año 2021, el comercio electrónico continuó creciendo de forma agigantada: fueron vendidas 381

millones de unidades, un 52% más que en el año anterior, a través de 196 millones de órdenes de compra. Además, se sumaron 684.459 nuevos compradores, registrando una suma total de 20.742.665 compradores online [1]. En este contexto, los portales de comercio electrónico como region20[1] se han convertido en una fuente fundamental para brindar al usuario una forma de poder comercializar sus productos (nuevos o usados) o servicios. En este tipo de portales, tanto usuarios particulares como comercios locales ven la posibilidad de acceder a esta forma de comercialización de productos o servicios a un costo reducido.

En la Figura 1 se puede apreciar la página de inicio del portal region20. En dicha página se destaca la posibilidad tanto de adquirir un producto así como también vender productos. Por lo anterior es que se distinguen dos tipos de roles que un usuario puede desenvolver dentro del portal: como comparador y como vendedor.

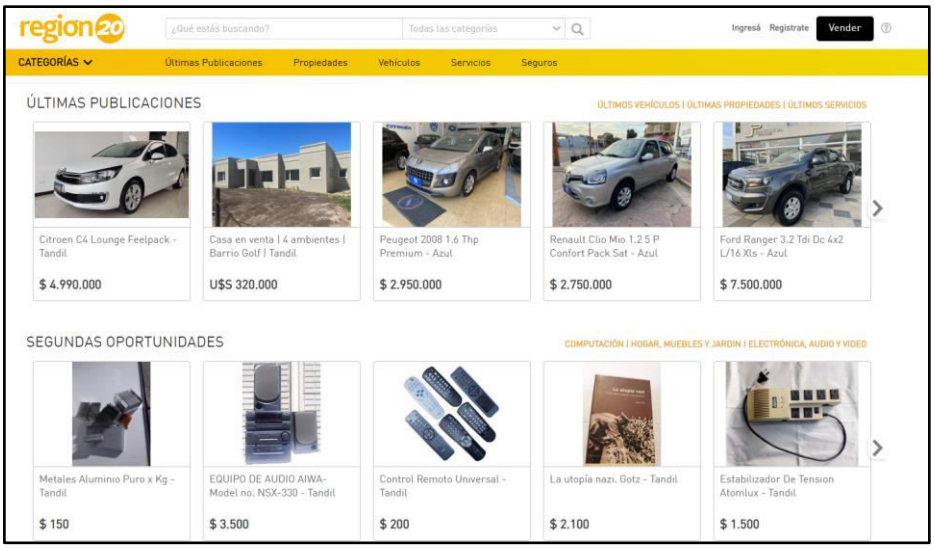


Figura 1 - Página de inicio del portal region20[1]

Para lograr una efectiva simbiosis entre el comprador y el vendedor es necesario poder llegar al comprador con los productos que éste requiere, en el tiempo que los requiere y con la posibilidad de acceder a éstos en el menor tiempo posible. Para poder realizar esto es necesario conocer el perfil del comprador, es decir tanto sus características generales que definen sus hábitos de compra, como así también sus necesidades puntuales en el corto plazo. Para la realización de una venta efectiva también debe superarse el umbral de espera, es decir la posibilidad de unir al comprador con el producto en el menor tiempo posible. La logística en la entrega de una venta es uno de los principales desafíos que enfrenta el comercio electrónico.

En este contexto, el proyecto abordado tuvo por objetivo realizar un prototipo que pueda acercar a las partes (comprador-vendedor) mediante una caracterización inteligente de cada usuario junto con sus preferencias a corto plazo. De esta forma, se

[1] https://www.region20.com.ar/

buscó poder brindar alertas inteligentes sobre artículos que puedan interesarles a los usuarios y, de ser posible, en zonas de proximidad geográficas. El proyecto consideró como aspectos fundamentales para su fácil adopción la posibilidad de adquirir el conocimiento del usuario de forma no intrusiva y con información pre-existente.

Este artículo surge como parte de los resultados de un proyecto Fase Cero financiado por la Fundación Sadosky entre dos institutos de investigación de la Facultad de Ciencias Exactas de la UNCPBA, ISISTAN (CONICET - UNCPBA) e INTIA, y la empresa Región Global. Esta empresa posee un sitio de comercio electrónico zonal denominado region20.

El artículo se encuentra organizado de la siguiente manera. En la sección 2 se mencionan los antecedentes del proyecto. En la sección 3 se detallan los objetivos. En la sección 4 se describen las etapas del proyecto. En la sección 5 se describen los resultados alcanzados. Finalmente en la sección 6 se exponen las conclusiones.

2 Trabajos relacionados

El problema del perfilado y clasificación de clientes ha sido estudiado anteriormente. Distintos enfoques han sido utilizados para el perfilado de usuarios en e-commerce utilizando información sobre la navegación y patrones de conducta de los clientes ya sea para clasificar a los clientes [2, 3 ,4] o para poder predecir su próxima compra [5]. En este contexto es importante identificar a los usuarios dentro del e-commerce para realizar acciones tales como la recomendación de anuncios y poder convertirlos de usuarios a clientes.

La recomendación de información contextualizada a partir de la geolocalización del usuario es un área que ha recibido especial atención en la comunidad de sistemas de recomendación en los últimos años. Este tipo de campañas publicitarias, complementan las campañas por medios tradicionales (televisión, radio, Internet) permitiendo la generación de campañas publicitarias personalizadas de acuerdo a los intereses del usuario, su ubicación actual, sus necesidades del momento y los dispositivos que utilizan.

En [6], por ejemplo, se presenta una infraestructura para la recomendación de publicidades de actividades tanto comerciales como no comerciales, buscando minimizar las interacciones del usuario e intentando aprender lo más posible acerca de sus hábitos de manera implícita. Esta arquitectura, hace uso de las antenas de telefonía celular para deducir la ubicación del usuario, con una precisión mucho inferior a la que podría obtenerse utilizando las antenas GPS. La información utilizada para caracterizar tanto a los usuarios como a los anuncios a recomendar incluyen las siguientes características: categoría (venta por mayor y menor, arte y entretenimiento, etc), día (días laborables, fines de semana), franja horaria (antes de las 17 hs y después de las 17 hs), lugar (exterior, interior y formal, interior e informal), costo (gratis, con costo), ejecutor (celebridades, otros). El modelo predictivo utiliza una red neuronal de dos capas y estereotipos de usuarios para recomendar una lista acotada y ordenada de anuncios a los usuarios.

En el trabajo presentado en [7] también se utiliza un aprendizaje de preferencias de los usuarios a partir de sus actividades online. Este enfoque, utiliza tanto la información de los productos por los que el usuario navega en un sitio de comercio electrónico, como los comentarios dejados por los clientes acerca de los productos adquiridos. Recientemente, en [8] se concluye que en la publicidad basada en la ubicación, el diseño publicitario más efectivo se basa en ofertas ordenadas por distancia. Los resultados muestran que la distancia geográfica tiene un impacto negativo en el comportamiento de elección de cupones de ofertas (es decir, cuanto más alejados están los consumidores de una tienda, menos probabilidades tienen de elegir un cupón relacionado), lo que puede explicarse por los costos de transporte. Sin embargo, también se observa que los consumidores aceptan mayores costos de transporte para recibir descuentos más atractivos.

El trabajo desarrollado en [9, 10, 11] obtiene automáticamente información de geolocalización de los incidentes de tránsito que afectan a una determinada región, a partir del conocimiento extraído de publicaciones en redes sociales. La solución ad-hoc desarrollada en este proyecto servirá de puntapié inicial para analizar e identificar información acerca de la ubicación de los productos o servicios ofrecidos en diferentes anuncios publicitarios cuando no se cuente explícitamente con esta información.

Relativo al aprendizaje de perfiles de rutinas de transporte de un usuario, en [12] se presentó un enfoque de un asistente inteligente para asistir a los ciudadanos en el traslado urbano. El prototipo desarrollado permite aprender de manera no intrusiva la rutina de un usuario y, a partir de esto, realizar sugerencias que le permitan evitar problemas en su traslado (congestionamientos, manifestaciones, accidentes, calles cortadas, etc). De manera similar, el enfoque presentado en [13] aprende los intereses de los usuarios relativos a atracciones turísticas. Los perfiles de usuario se infieren automáticamente en base a los recorridos realizados por los usuarios y el tiempo de permanencia en cada punto de interés (POI). En base a esta información histórica, cuando un usuario visita una nueva ciudad, se le ofrecen atracciones que pueden serle de interés.

El portal regional region20 fue lanzado al mercado en 2006 por la empresa Región Global S. A., y cuenta actualmente con más de 150,000 usuarios registrados. El marketplace no cobra comisión por ventas y tiene la posibilidad de publicar artículos gratis lo que hace que exista un flujo muy importante de publicaciones.

Como en cualquier marketplace, los usuarios registrados en el portal tienen la posibilidad de publicar anuncios de venta/alquiler de productos y servicios. Estos anuncios se organizan inicialmente en cuatro grandes ejes, “Vehículos”, “Propiedades”, “Artículos” y “Servicios”, que luego son subdivididos en numerosas categorías y subcategorías.

Como su nombre lo indica, el eje de “Vehículos” agrupa anuncios de todo tipo de rodados. La clasificación inicial se realiza en “Autos”, “Camiones”, “Camionetas”, “Cuatriciclos”, “Motos”, “Utilitarios” y “Otros Vehículos”. que luego son agrupados según marca y modelo. El contenido del filtro de “marca” va a depender del “tipo” de vehículo seleccionado, así como el “modelo” dependerá de la “marca” previamente seleccionada.

El eje “Propiedades” agrupa anuncios de diferentes tipos de inmuebles, tales como casas, departamentos, locales, lotes, negocios, galpones, entre otros, disponibles tanto para la venta como para alquiler.

El eje “Artículos” incluye a los anuncios generales de venta de productos (que no sean propiedades o vehículos). Similar a los anteriores, el artículo se puede clasificar en categorías predefinidas de hasta 3 niveles de profundidad. Por ejemplo, la categoría de nivel 1 “Alimentos y Bebidas”, posee como subcategorías de nivel 2 “Accesorios”, “Almacén”, “Bebidas”, “Comidas preparadas”, “Dulces y postres”, “Frescos”, y “Otros”. Dentro de “Bebidas”, por ejemplo, se encuentran las categorías de nivel 3 “Aguas”, “Bebidas blancas”, “Cervezas”, “Jugos y Gaseosas”, “Licores”, “Vinos” y “Whiskies”. Puede ser que una categoría no posea los tres niveles de profundidad, por ejemplo “Art. Regionales y regalos”, es una categoría con un solo nivel.

Por último, los anuncios pertenecientes al eje de “Servicios”, a diferencia de los anteriores, no representan la venta de un producto físico. Estos se clasifican en las categorías “Belleza y Cuidado personal”, “Imprenta”, “Profesionales”, “Transporte General”, “Clases y Capacitaciones”, “Mantenimiento de Vehículos”, “Ropa y Moda”, y “Turismo y recreación”, cada uno con diferentes subcategorías, según corresponda.

En los 4 tipos de avisos que se pueden crear en el portal, se solicita un título, descripción, (texto libre), precio (moneda y monto), forma de pago y la posibilidad de cargar imágenes. En el caso de vehículos y propiedades se solicitan además datos estructurados para una mejor caracterización de los mismos. Algo similar ocurre en servicios donde se busca visibilizar características del servicio prestado.

En la actualidad, para publicar un anuncio, los usuarios deben seleccionar al inicio de la publicación y de forma manual, una categoría de las ofrecidas por el portal, respetando la jerarquía de categorías preestablecida. En la Figura 2 se muestra un ejemplo de una nueva publicación dentro del eje Artículos. Luego, tres moderadores se encargan de revisar manualmente cada uno de los anuncios publicados controlando, entre otras cosas, que se encuentren bien categorizados. El proceso de selección de una categoría puede resultar tedioso y confuso para los usuarios ya que las mismas al estar organizadas en categorías jerárquicas (en el peor de los casos en cuatro niveles que se van desplegando al elegir una categoría padre) existe cierta ambigüedad y superposición entre alguna de ellas. Es por esto que fue posible identificar en el sitio, una gran cantidad de anuncios mal clasificados por los usuarios, o inconsistencias entre las categorías de ciertos anuncios (por ejemplo, diferentes anuncios de un mismo modelo de automotor que se encontraba clasificados como autos, camionetas o utilitarios de forma aleatoria). Algo difícil de detectar incluso para los moderadores ya que se le presenta un único aviso y debe decidir si está correctamente clasificado, y al verlos individualmente sin contexto puede parecer que todos están bien clasificados. Automatizar por lo tanto la selección de la categoría de un anuncio tiene un impacto directo tanto sobre la usabilidad del sitio como en la mejora del trabajo de los moderadores.

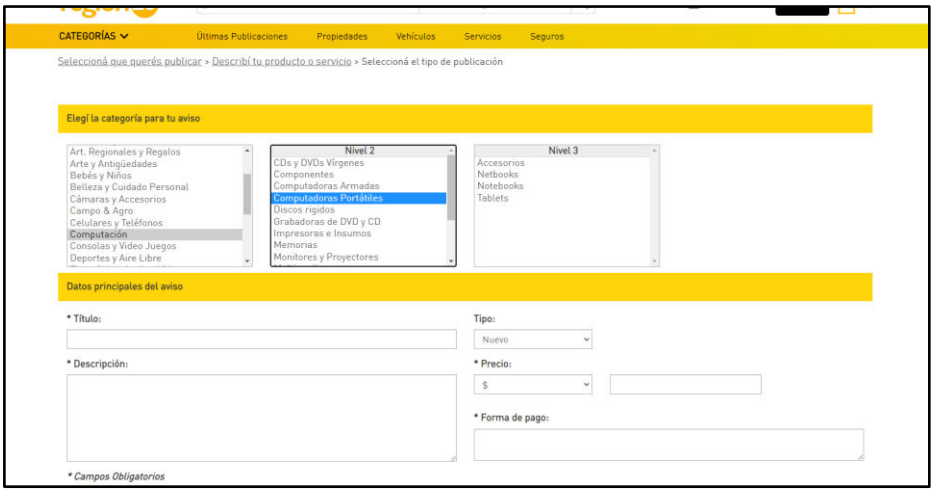


Figura 2 - Publicación de un producto dentro de la categoría Artículos

3 Objetivos del proyecto

El objetivo general del proyecto consistió en diseñar un prototipo de asistente inteligente que permitiera vincular las necesidades de un usuario (capturadas de forma implícita) con las ofertas del portal de comercio electrónico region20, ordenando las mismas por proximidad geográfica al usuario.

Los objetivos específicos del mismo consistieron en:

1. Definir indicadores a extraer a partir de la información generada por el usuario en la navegación/búsqueda en el portal, e implementar algoritmo de extracción de dicha información.
2. Definir e implementar un algoritmo de extracción de indicadores a partir de las publicaciones de productos / servicios y los usuarios que las realizan, poniendo especial énfasis en aspectos geoespaciales.
3. Definir e implementar un algoritmo de perfilado de usuarios que permita, de forma inteligente, caracterizar sus necesidades en el mediano y corto plazo, a partir de los indicadores extraídos.
4. Definir e implementar métricas de similitud de usuarios que permitan minimizar los problemas del "arranque en frío" del algoritmo de predicción.
5. Definir e implementar un prototipo asistente inteligente que permita recomendar productos a los usuarios teniendo en cuenta sus necesidades y restricciones geoespaciales.
6. Definir un framework de evaluación que permita medir el éxito en las recomendaciones.

4 Etapas del proyecto

Para poder llevar a cabo el proyecto, el mismo se dividió en tres etapas que buscaban abarcar los objetivos anteriores:

Etapa 1: Extracción del conocimiento

En esta etapa se abordaron los objetivos específicos 1 y 2: para este fin fue necesario conocer la información contextual disponible.

- Documentación de la información disponible en el portal region20 y posibles formas de acceso a la misma.
- Definición de metodología de trabajo.
- Análisis de factibilidad de los algoritmos existentes en la literatura.
- Identificación de técnicas de recomendación utilizadas en el dominio.
- Selección de indicadores aptos y algoritmo de extracción de perfiles de usuario y productos.

Como resultado, se elaboró un informe técnico con un resumen detallado de la información disponible para la caracterización de los usuarios y de las publicaciones, un análisis de los algoritmos de perfilado de usuario existentes y un listado de aquellos que resulten más aptos para el problema.

En el informe mencionado se realizó un análisis de los datos disponibles dentro del portal region20 que pudieran dar lugar a la generación de recomendaciones de anuncios para los usuarios. Dicho análisis se realizó en conjunto con todos los integrantes del proyecto a lo largo de las diferentes reuniones, ya que no sólo se consideraron los datos disponibles sino también se analizó la utilidad de los mismos para el objetivo del proyecto. Asimismo, se realizó un análisis de las técnicas a considerar para la realización de un sistema de recomendación. Además de la descripción de las técnicas se incluyeron detalles respecto a la aplicación de las mismas en el dominio y, particularmente, en el portal region20. Por ejemplo, el filtrado colaborativo no resultaba ser una técnica de aplicación directa debido a que en region20 no se dispone de un rating de los artículos por parte de los usuarios. Sin embargo, cabía la posibilidad de realizar adaptaciones para poder inferir esos ratings a partir de datos de visitas, solicitud de datos y/o preguntas realizadas. Otra diferencia que podemos resaltar del dominio abordado con respecto a otros dominios en los que habitualmente se aplican los sistemas de recomendación es la volatilidad de los ítems. A diferencia de dominios como películas o libros, en el comercio electrónico y particularmente en sitios de venta como region20, los ítems están disponibles por un determinado tiempo hasta que son comprados y por lo tanto es de mayor relevancia que el modelo aprenda los intereses de los usuarios en cierto tipo de ítems que en un ítem en particular. El dominio ofrece, por otro lado, la posibilidad de explorar "ítems relacionados" a un producto en el cual un usuario mostró interés. Por ejemplo, si el usuario demostró interés en un teléfono celular, en lugar de recomendarle otros teléfonos celulares sería también válido ofrecerle accesorios tales como fundas, vidrios templados, auriculares, etc. En general, para cada una de las técnicas se consideró la información disponible en el dominio y se realizó un análisis de factibilidad.

Del trabajo en conjunto realizado en cada uno de los encuentros semanales, se pudieron también establecer diferentes puntos de uso de las recomendaciones, como por ejemplo un mail personalizado, en lugar de los actuales mails generales con una

newsletter, así como también adaptaciones a realizar al sitio para poder enriquecer las técnicas de recomendación. Por ejemplo, al momento de inicio del proyecto los avisos visitados por un usuario se guardaban solo por 15 días, debido a que el portal considera que luego de 15 días el interés por los avisos es muy probable que haya cambiado. Sin embargo, para poder consolidar los perfiles de los usuarios sería deseable contar con un periodo de tiempo más prolongado de información, siempre considerando el factor temporal y el cambio de preferencias en su consolidación. Otra posibilidad con la que resultaba interesante contar para mejorar los perfiles era la posibilidad de registrar los contactos entre las partes compradora y vendedora por medio de WhatsApp de manera de tener otra fuente de información para enriquecer los perfiles de los usuarios del sistema. Ambos ítems fueron atendidos por la empresa e incorporados al sitio a fin de poder trabajar con mayor volumen de información.

Etapa 2: Perfilado inteligente del Usuario

En esta etapa se diseñaron e implementaron los perfiles de usuario, se utilizó un algoritmo de caracterización individual del usuario y métricas de similitud de usuarios que permitieron realizar recomendaciones a usuarios sin suficiente información individual. Esta etapa abarcó los objetivos específicos 3 y 4.

- Análisis de técnicas de perfilado individual de un usuario para la recomendación de productos. Se puso un énfasis especial en la información general y en la de corto plazo, es decir interés actual sobre una compra.
- Análisis de métricas de similitud entre usuarios. Sobre la base de los indicadores individuales se realizó un análisis de técnicas de clustering que permitan una mejor caracterización del usuario nuevo del sistema.
- Diseño de algoritmo de perfilado inteligente de usuario. A partir de los análisis previos, diseñar un algoritmo que permita perfilar a un usuario a partir de la información disponible.
- Implementación del prototipo del perfilador inteligente de usuarios .

Como resultado, se elaboró un informe y se desarrolló el prototipo de algoritmo perfilador de usuarios.

En este informe se detalló el trabajo realizado en la segunda etapa del proyecto, en la cual se trabajó en la construcción del perfil de usuario. Para ello, en primer lugar se analizaron de manera conjunta los datos disponibles. Luego, se presentaron alternativas para construir el perfil de usuario, y en función de lo que se definió se evaluaron diferentes técnicas. Dado que parte de los datos se encuentran en formato textual, se describieron técnicas de Procesamiento de Lenguaje Natural. Finalmente, se analizaron y compararon diferentes técnicas de clasificación, que permiten asociar automáticamente una categoría a un artículo a partir de su título o su descripción. El trabajo en conjunto entre la empresa y los institutos permitió que a lo largo de esta etapa se fueran aclarando las diferentes implicancias de cada una de las partes dentro del ecosistema de la empresa y su impacto.

Un perfil es una descripción de alguien que contiene lo más importante o interesante sobre él o ella en un dominio particular. En nuestro caso en particular, el perfil o modelo de usuario contiene información esencial sobre un usuario individual para poder proveer recomendaciones de potenciales productos de interés. La motivación de crear

perfiles de usuario es que los usuarios difieren en sus preferencias, intereses, antecedentes y objetivos al utilizar aplicaciones de software, y descubrir estas diferencias es vital para proporcionar a los usuarios servicios personalizados.

Para construir el perfil de usuario se tomaron las diferentes fuentes de datos disponibles que representan interacciones de usuarios con artículos:

- la información de las visitas del usuario a diferentes avisos,
- las búsquedas realizadas por el usuario,
- el historial de solicitudes,
- las preguntas realizadas por el usuario.

Se pudo realizar un consenso respecto a la importancia de dos ejes con los cuales medir el usuario respecto a las interacciones con el sitio: **interacciones directas e indirectas**. Las interacciones directas surgen de la interacción del usuario con elementos puntuales y cuantificables del sitio, en nuestro caso particular con los avisos o productos. Estas interacciones las consideramos directas ya que se posee información explícita de los avisos y por ende sus categorías e información específica. Existe una forma adicional de interacción que realiza el usuario, la cual se considera indirecta, y es cuando el usuario realiza una búsqueda. En este caso, el usuario ingresa ciertas palabras claves que considera importantes y que por ende lo van a llevar a una posible interacción directa de las mencionadas en la sección anterior. En este punto, es importante realizar un registro de este tipo de interacciones (Listado de las palabras de búsqueda) pero asociándolo a categorías potenciales de interés del usuario.

Para la construcción de las interacciones indirectas se procedió a inferir qué categorías eran las de interés para el usuario a partir de los términos utilizados para las búsquedas realizadas, y se registraron las 3 primeras categorías según su orden de importancia relativa a la búsqueda. Para esto fue necesario procesar el texto de la búsqueda y llevar esa información a una asociación con las categorías existentes en el sistema.

Sin embargo, cuando se desea procesar computacionalmente textos escritos de forma libre (un campo de texto en el cual un usuario escribe sentencias sin estructura preestablecida), estos deben primero convertirse a un formato que sea entendible por la computadora y apto para la realización de cálculos. El formato más aceptado para esta tarea es conocido como “bags of words”, o bolsa de palabras, cuya representación computacional es a través de vectores. Para poder llegar a la representación de cualquier texto en los modelos, es necesario un procesamiento previo que aplica un conjunto de tareas, normalmente secuenciales, que se describen a continuación.

- Tokenización y segmentación: consiste en dividir el texto en entidades significativas llamadas tokens, En nuestro caso particular los signos de puntuación no serán utilizados así como tampoco vamos a distinguir entre mayúsculas y minúsculas..
- Normalización: se refiere a una serie de tareas relacionadas destinadas a poner todo el texto en igualdad de condiciones: convirtiendo todo el texto en mayúsculas o minúsculas, eliminando la puntuación, convirtiendo los números a sus equivalentes de palabras.Este proceso en nuestro caso (y por lo

general para cualquier dominio) está muy apegado al realizado de forma estándar.

- Eliminación de stop-words: Las "stop-words" o "palabras vacías" son las palabras en cualquier idioma que no agregan mucho significado a una oración ya sea por su frecuencia o por su semántica. En nuestro caso particular además de las palabras habituales, los trabajos exploratorios realizados permitieron descubrir stop-words que no aportan información al problema atacado. Por ejemplo "hola"y "gracias" (por citar dos emblemáticas) son palabras que no nos permiten determinar características distintivas del problema.
- POS-Tagging: part-of-speech tagging o etiquetado gramatical consiste en clasificar y etiquetar las palabras dentro de un texto según su significado gramatical tal como verbo, sustantivo, adjetivo, preposición. En nuestro caso no se realizó un pos-tagging ya que no tenía una implicancia sobre el problema a resolver.

Lo anterior se aplicó tanto a los títulos como a las descripciones de las publicaciones de cada categoría del sistema, dando como resultado un conjunto de palabras que describen a cada categoría. En la Figura 3 se muestra, en formato de nube de palabras, los términos que describen a la categoría computación, calculadas a partir del procesamiento de los títulos y las descripciones de las publicaciones dentro de la categoría.

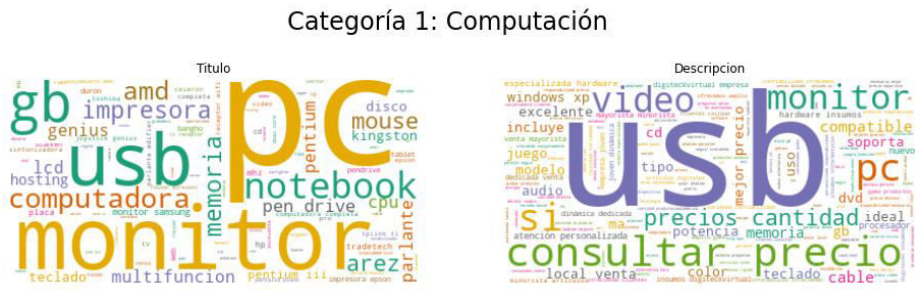


Figura 3 - Nube de palabras para categoría computación. A mayor tamaño de la palabra, mayor la importancia de la misma dentro de la categoría.

Como una primera aproximación a la clasificación de avisos en categorías, en esta etapa se exploró si era viable predecir, dado el título o la descripción de un artículo determinado, la categoría asociada. En la plataforma existen diferentes categorías pre-establecidas y son las que el usuario selecciona para poder caracterizar de mejor manera a su aviso. Algunos ejemplos de categorías son: Libros, Cámaras y accesorios, Juegos, Deportes, Accesorios de autos, Vehículos, Propiedades. Cuando un usuario publica un aviso, lo asocia manualmente a una única categoría, pero estas categorías están organizadas en hasta 4 niveles de profundidad.

Las categorías de la base de datos utilizada para el presente proyecto, detalla 858 categorías diferentes. Debido a esto se llegó a la conclusión de que una clasificación directa en un número tan elevado de clases difícilmente lleve a resultados aceptables. En una primera etapa del proyecto, se limitó el trabajo a las categorías de 1 Nivel, es decir a las categorías que en la base de datos no poseen un padre. El resultado es un total de 28 categorías a utilizar como clases en el proceso de clasificación de los artículos.

Se probaron diferentes modelos de clasificación que permitieron determinar la categoría a partir de las palabras utilizadas en sus artículos (ya sea título o descripción) La Figura 4 muestra los resultados de la comparación de cuatro modelos de clasificación para el título de los artículos. Se puede apreciar que se obtuvo una precisión media de entre el 80 y el 85%.

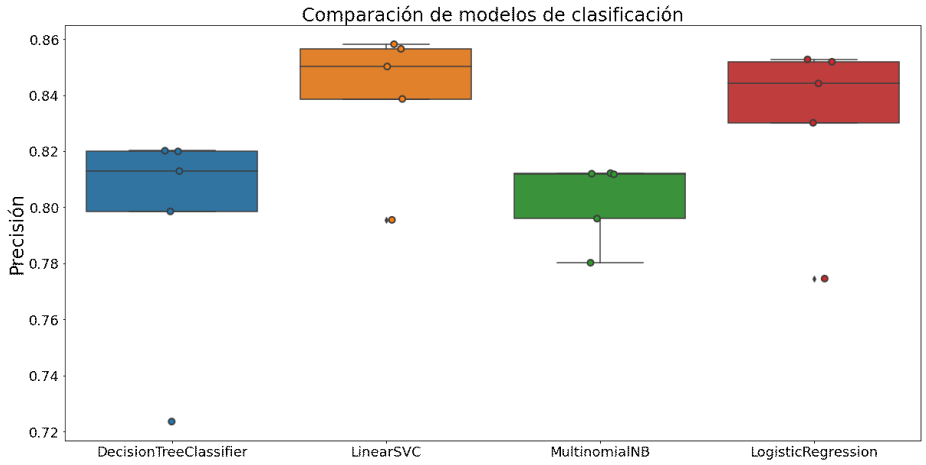


Figura 4 - Boxplot que ilustra la performance de cuatro modelos de clasificación aplicados sobre los títulos de los artículos

La precisión media, sin embargo, suele ser una métrica que si bien da una idea global de la performance de un modelo de clasificación, puede resultar engañosa en situaciones como la que estamos abordando: conjuntos de datos altamente desbalanceados en su distribución de clases. Es por esto que a continuación exploramos la matriz de confusión obtenida mediante una validación utilizando la técnica de Hold-out, utilizando un 66% de los datos para entrenar el modelo de clasificación y el 33% restante para testeo. Para esta evaluación, utilizamos el modelo LinearSVC que mostró mejor performance en los resultados previos. La Figura 5 muestra la matriz de confusión obtenida. A primera vista se puede observar que, a pesar del desbalance de clases, la clasificación para la mayoría de las categorías es buena. A modo de ejemplo, para la categoría “Vehículos” que representa la clase mayoritaria, pudieron identificarse correctamente un total de 54.080 artículos, que representa un 98% del total de los artículos pertenecientes a esa categoría (recall). La precisión de esta clase (artículos

para los que se predijo correctamente la categoría) resultó del 95%. Las clases con peor performance, fueron “Salud y Equipamiento Médico”, la clase minoritaria, con un 67% de precisión y solo un 16% de recall, y la clase “Otros”, con 39% de precisión y 19% de recall. En la Figura 6 se muestra el detalle de los valores de las métricas obtenidas para cada categoría, junto a la precisión macro promedio y la precisión pesada por cantidad de artículos pertenecientes a cada categoría.

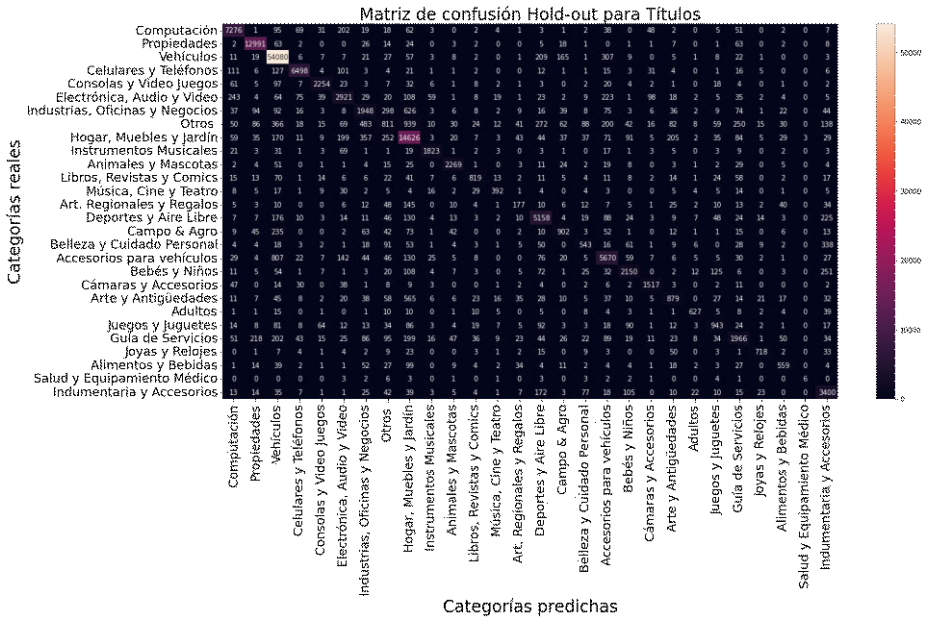


Figura 5- Matriz de confusión obtenida mediante la técnica de Hold-out con un 66% de datos de entrenamiento utilizando los títulos de los avisos

	precision	recall	f1-score	support
Computación	0.90	0.92	0.91	7942
Propiedades	0.96	0.98	0.97	13234
Vehículos	0.95	0.98	0.97	54980
Celulares y Teléfonos	0.95	0.93	0.94	6974
Consolas y Video Juegos	0.90	0.88	0.89	2564
Electrónica, Audio y Video	0.75	0.73	0.74	4018
Industrias, Oficinas y Negocios	0.60	0.56	0.58	3485
Otros	0.39	0.19	0.26	4216
Hogar, Muebles y Jardín	0.80	0.89	0.84	16471
Instrumentos Musicales	0.91	0.90	0.91	2025
Animales y Mascotas	0.91	0.91	0.91	2484
Libros, Revistas y Comics	0.79	0.69	0.74	1192
Música, Cine y Teatro	0.79	0.69	0.74	566
Art. Regionales y Regalos	0.43	0.30	0.35	588
Deportes y Aire Libre	0.81	0.85	0.83	6062
Campo & Agro	0.67	0.59	0.63	1530
Belleza y Cuidado Personal	0.61	0.43	0.50	1271
Accesorios para vehículos	0.80	0.79	0.80	7182
Bebés y Niños	0.79	0.74	0.77	2908
Cámaras y Accesorios	0.86	0.89	0.88	1702
Arte y Antigüedades	0.61	0.46	0.52	1925
Adultos	0.88	0.82	0.85	762
Juegos y Juguetes	0.68	0.60	0.64	1567
Guía de Servicios	0.68	0.58	0.63	3392
Joyas y Relojes	0.87	0.81	0.84	891
Alimentos y Bebidas	0.71	0.60	0.65	924
Salud y Equipamiento Médico	0.67	0.16	0.26	38
Indumentaria y Accesorios	0.72	0.84	0.78	4052
accuracy			0.86	154945
macro avg	0.76	0.70	0.72	154945
weighted avg	0.85	0.86	0.86	154945

Figura 6 - Detalle de los resultados obtenidos a partir de la matriz de confusión presentada en la Figura 5.

Etapa 3: Recomendación de Productos

A partir del trabajo incremental en esta etapa se trabajó fuertemente sobre la disparidad y diversidad importante dentro de las categorías del sistema. Punto importante a tener en cuenta ya que eran más de 800 categorías a predecir y los resultados experimentales obtenidos no eran buenos para trabajar sobre ellas individualmente ya que muchas poseían muy pocos artículos. Es importante notar que las categorías de las publicaciones se organizan en jerarquías, Esto último es importante

ya que este conocimiento no se aprovecha si se aplana toda le estructura y se clasifica directo sobre la categoría final de un anuncio. En esta etapa surgió la necesidad de trabajar con un modelo de clasificadores concatenados o meta clasificador, el mismo se presenta en la Figura 7. En la figura se puede apreciar cómo a partir del título, se ingresa al primer clasificador que determina una de las 4 categorías principales (o ejes verticales), es decir Artículo, Vehículo, Propiedad o Servicio. Este primer clasificador se entrena con todas las publicaciones del sitio pero se utiliza como variable objetivo (clase) la categoría de primer nivel asociada a la categoría real del anuncio.

Con el Título y la primera categoría predicha, se prosigue a los clasificadores del Nivel 1. En este meta clasificador coexisten 4 clasificadores distintos, uno por cada una de las posibles 4 categorías del primer nivel. Estos clasificadores determinan las categorías en las que se subdivide la categoría dada, por ejemplo en Artículos este clasificador abarca categorías que van desde Computación hasta Salud. Para entrenar cada uno de estos clasificadores, el conjunto de anuncios original se subdivide en cuatro subconjuntos, de acuerdo a la categoría principal de cada anuncio. Este procedimiento se repite un nivel más (siempre que la sub categoría se subdivida nuevamente, en caso contrario se llegó a una hoja del árbol de categorías y se da por finalizada la clasificación). Por ejemplo, si un aviso se clasifica como *Artículo* y luego *Computación*, se usa el clasificador del Nivel 2 para determinar la subcategoría dentro de *Computación*, por ejemplo *Periféricos*.

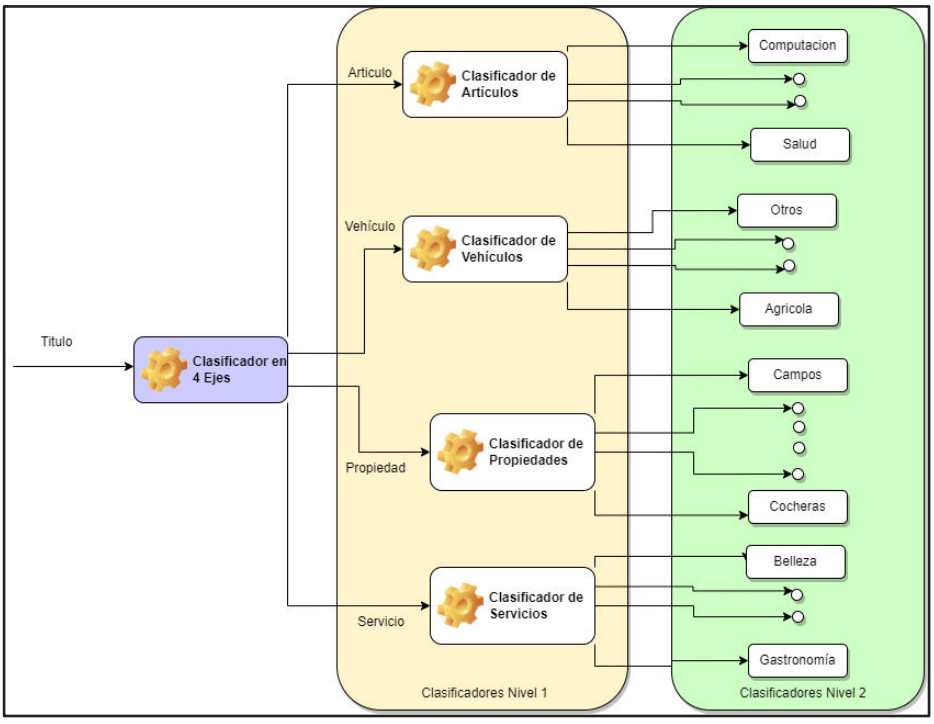
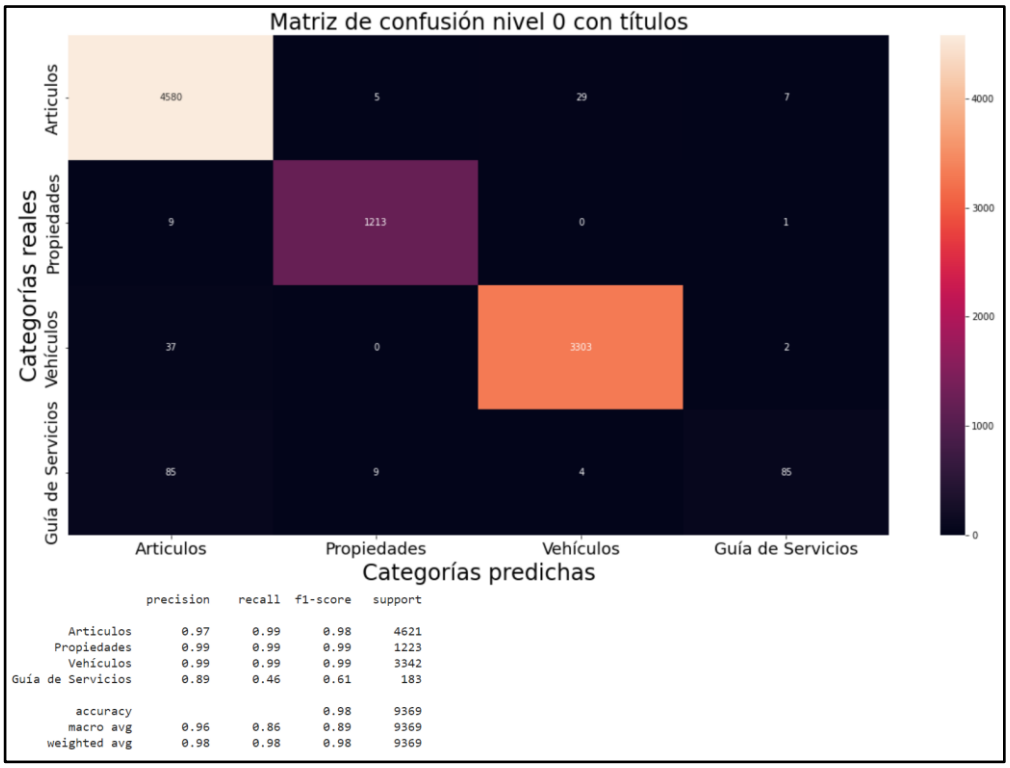


Figura 7 - Clasificador de categorías concatenados.

La Figura 8 muestra los resultados de la matriz de confusión resultante de la clasificación en los cuatro ejes principales del portal de cada artículo utilizado para prueba, junto a los valores de las métricas precisión, recall y f1-score. Como puede apreciarse, para las categorías *Artículos*, *Propiedades* y *Vehículos*, la clasificación es casi perfecta. Las mayores equivocaciones del algoritmo de clasificación se obtienen para la categoría *Guía de Servicios*, lo cual es esperable por ser la clase de la que se disponen menos anuncios para entrenamiento. De todas formas, la precisión obtenida para esta clase es alta (89%), resultando el recall notablemente inferior (46%). Esto quiere decir que, los anuncios para los que el algoritmo predice la categoría *Guía de Servicios* se corresponden con artículos de esta categoría en el 89% de los casos, pero que solo se encontró el 46% de anuncios pertenecientes a la categoría. Observando la matriz de confusión podemos ver que la mayor parte de estos anuncios fueron clasificados como *Artículos*. Explorando algunos de los títulos correspondientes a estos anuncios mal clasificados (Figura 9) puede observarse que incluso para el razonamiento humano resulta difícil clasificar, a partir de las palabras resultantes, la categoría *Guía de Servicios*. En las reuniones de coordinación los integrantes de la empresa manifestaron que en muchos de estos casos se determinaba que era un servicio más por el usuario que publicaba el anuncio que por el título del mismo. Es necesario recordar que los anuncios son moderados por un administrador encargado de verificar las categorías asignadas por el usuario antes de que el anuncio mismo sea publicado.



	precision	recall	f1-score	support
Articulos	0.97	0.99	0.98	4621
Propiedades	0.99	0.99	0.99	1223
Vehículos	0.99	0.99	0.99	3342
Guía de Servicios	0.89	0.46	0.61	183
accuracy			0.98	9369
macro avg	0.96	0.86	0.89	9369
weighted avg	0.98	0.98	0.98	9369

Figura 8 - Matriz de confusión y métricas obtenidas para la clasificación en los 4 ejes principales del portal

```
Articulo con id  473948
Titulo:  Plastine. Papel para hornear, papel
Categoría real 195 (Guía de Servicios)
Categoría predicha -1 (Articulos)
El titulo del documento 52 tiene las palabras resultantes
hornear (1)
papel (2)
plastine (1)
========================================================================
Articulo con id  457661
Titulo:  Plotter para guardapolvos en vinilo
Categoría real 195 (Guía de Servicios)
Categoría predicha -1 (Articulos)
El titulo del documento 478 tiene las palabras resultantes
plotter (1)
vinilo (1)
========================================================================
Articulo con id  474917
Titulo:  Botines Nike Total 90
Categoría real 195 (Guía de Servicios)
Categoría predicha -1 (Articulos)
El titulo del documento 712 tiene las palabras resultantes
botines (1)
nike (1)
total (1)
```

```
======================================================================
Articulo con id  471273
Titulo:  ELEGÍ HACER LA CONVERSIÓN A GNC EN OLAVA
Categoría real 195 (Guía de Servicios)
Categoría predicha -1 (Articulos)
El titulo del documento 781 tiene las palabras resultantes
gnc (1)
======================================================================
Articulo con id  455052
Titulo:  Parrilla "El Recreo" - Parrillaelrecreo
Categoría real 195 (Guía de Servicios)
Categoría predicha -1 (Articulos)
El titulo del documento 903 tiene las palabras resultantes
parrilla (1)
```

Figura 9 - Ejemplo de 5 de los 85 anuncios de artículos pertenecientes a la categoría Guía de Servicios y clasificados como artículos

Una vez determinada la categoría del primer nivel se utilizaron los clasificadores correspondientes a las categorías de segundo nivel. A continuación se muestran las métricas obtenidas para cada uno de los 4 clasificadores específicos utilizados.

	precision	recall	f1-score	support
Casas	1.00	0.99	1.00	406
Departamentos	1.00	1.00	1.00	462
Locales	1.00	1.00	1.00	40
Negocios	1.00	1.00	1.00	11
Quintas	1.00	1.00	1.00	26
Lote	0.99	1.00	1.00	214
Campos	1.00	1.00	1.00	22
Oficinas	1.00	1.00	1.00	14
Galpónes	1.00	1.00	1.00	15
Cocheras	1.00	1.00	1.00	12
accuracy			1.00	1222
macro avg	1.00	1.00	1.00	1222
weighted avg	1.00	1.00	1.00	1222

Figura 10 - Métricas obtenidas para la categoría Propiedades

	precision	recall	f1-score	support
Belleza y cuidado personal	0.77	0.91	0.83	11
Clases y Capacitaciones	0.70	0.54	0.61	13
Fiestas y Eventos	0.00	0.00	0.00	2
Mantenimiento de Vehículos	1.00	0.14	0.25	7
Profesionales	0.88	0.29	0.44	24
Turismo y Recreación	1.00	0.67	0.80	3
Mantenimiento para el Hogar	0.72	0.82	0.77	28
Servicio Técnico	0.80	0.67	0.73	12
Transporte General	0.75	0.60	0.67	10
Otros Servicios	0.52	0.83	0.64	59
Gastronomía	1.00	0.50	0.67	8
Imprenta	0.00	0.00	0.00	3
Ropa y Moda	0.00	0.00	0.00	3
accuracy			0.64	183
macro avg	0.63	0.46	0.49	183
weighted avg	0.68	0.64	0.61	183

Figura 11- Métricas obtenidas para la categoría Guía de Servicios

	precision	recall	f1-score	support
Autos	0.93	0.94	0.94	2114
Motos	0.93	0.97	0.95	207
Camiones	0.84	0.81	0.83	47
Camionetas	0.86	0.85	0.85	813
Otros vehiculos	0.85	0.76	0.80	62
Utilitarios	0.41	0.44	0.42	71
Cuatriciclos	0.89	0.61	0.72	28
accuracy			0.90	3342
macro avg	0.82	0.77	0.79	3342
weighted avg	0.90	0.90	0.90	3342

Figura 12 - Métricas obtenidas para la categoría Vehículos

	precision	recall	f1-score	support
Computación	0.92	0.88	0.90	706
Celulares y Teléfonos	0.93	0.91	0.92	246
Consolas y Video Juegos	0.90	0.82	0.86	145
Electrónica, Audio y Video	0.81	0.82	0.81	339
Industrias, Oficinas y Negocios	0.70	0.68	0.69	237
Otros	0.44	0.29	0.35	216
Hogar, Muebles y Jardín	0.71	0.89	0.79	964
Instrumentos Musicales	0.84	0.67	0.74	72
Animales y Mascotas	0.86	0.76	0.81	74
Libros, Revistas y Comics	0.82	0.61	0.70	133
Música, Cine y Teatro	0.43	0.43	0.43	7
Art. Regionales y Regalos	0.38	0.30	0.33	20
Deportes y Aire Libre	0.82	0.81	0.82	356
Campo & Agro	0.81	0.73	0.77	63
Belleza y Cuidado Personal	0.81	0.71	0.76	90
Accesorios para vehículos	0.80	0.82	0.81	234
Bebés y Niños	0.68	0.71	0.69	185
Cámaras y Accesorios	0.85	0.85	0.85	61
Arte y Antigüedades	0.53	0.47	0.50	99
Adultos	0.95	0.91	0.93	22
Juegos y Juguetes	0.54	0.33	0.41	58
Joyas y Relojes	0.88	0.58	0.70	12
Alimentos y Bebidas	0.94	0.74	0.82	102
Salud y Equipamiento Médico	0.60	0.55	0.57	22
Indumentaria y Accesorios	0.73	0.77	0.75	158
accuracy			0.78	4621
macro avg	0.75	0.68	0.71	4621
weighted avg	0.78	0.78	0.77	4621

Figura 13 - Métricas obtenidas para la categoría Artículos

El trabajo se siguió dividiendo acorde a lo estipulado anteriormente. El meta clasificador final obtenido toma un título, predice la categoría en primer nivel y con esa categoría predicha se selecciona el clasificador del siguiente nivel, al cual se le vuelve a pasar el título para predecir la subcategoría. Este proceso se repite hasta llegar a una hoja, es decir, una categoría que no tenga subcategorías.

5 Resultados alcanzados

A partir del desarrollo del proyecto se fueron alcanzando diversos objetivos planteados originalmente en el proyecto y otros nuevos que surgieron de la colaboración entre las partes del proyecto.

Perfil del usuario

Con el análisis de la información disponible en el portal, se estableció un perfil individual para cada usuario. Dicho perfil puede actualizarse a medida que se establecen nuevas interacciones. El perfil registra las interacciones directas e indirectas.

A partir del clasificador desarrollado también se registran las potenciales categorías de publicaciones en las que estaba interesado el usuario en el momento de hacer una búsqueda. La jerarquización de las características permite desarrollar futuras aplicaciones que permitan llegar de forma personalizada al usuario. El perfil permite:

- Obtener el interés histórico de un usuario
- Obtener el interés de un usuario dada una ventana de tiempo
- Dada una publicación, determinar si la misma resulta de interés o no a un usuario, con un grado de certeza.

Clasificador de publicaciones

Se propuso un enfoque de un meta clasificador de 3 niveles que permite obtener las posibles categorías a partir de un conjunto de palabras. Este clasificador, el cual fue evaluado y probado exhaustivamente, permite clasificar las interacciones de un usuario cuando realiza búsquedas en el portal, así como también, dado el título de una nueva publicación, categorizarla de manera automática. Este último aspecto resultó de suma importancia para la empresa, ya que colabora en minimizar tanto la redundancia como los errores humanos en la asignación de categorías.

Aplicaciones dentro del dominio

Actualmente, la empresa está finalizando con el proceso de migración del portal region20 a PHP 8. El próximo hito es la mejora en la publicación de avisos, donde se planea la integración del categorizador presentado y más adelante la obtención del perfil de usuarios. A partir de las diversas reuniones surgieron diferentes potenciales aplicaciones del proyecto realizado en colaboraciones futuras, por ejemplo:

- **Mailing personalizado**. El portal realiza campañas de mailing esporádicas con los últimos avisos o acorde a fechas importantes, tales como día del niño, día de la madre, etc. A partir de los objetivos alcanzados, es posible armar el mail de forma personalizada, es decir, dado un usuario incluir una serie potencial de artículos que le resulten de su interés primario.
- **Sugerencia de compradores para nuevos artículos publicados**. Cuando un nuevo artículo es aprobado para su publicación es posible obtener un listado de usuarios potencialmente interesados en el mismo.
- **Carrusel personalizado de artículos**. Actualmente, en la navegación del sitio se muestran artículos relacionados al artículo que el usuario está viendo y no artículos de su potencial interés. A partir de los resultados alcanzados en la ejecución del proyecto, será posible que dicho carrusel de artículos sea armado en base a los intereses de los usuarios.
- **Categorización de nuevos artículos**. En este caso es una funcionalidad que se encuentra en desarrollo y utiliza como pilar el clasificador desarrollado. Esta funcionalidad permite, dado un título de un anuncio a publicar, sugerir automáticamente las categorías más probables para asociar al mismo. La categorización de la predicción en tres etapas permite que la funcionalidad se integre de forma natural, ya que al usuario se le seleccionarán de forma automática las categorías predichas, pudiendo cambiarlas si lo desea. Este

clasificador permitirá establecer un criterio común entre los moderadores, lo que hará que con el tiempo todas las publicaciones converjan de forma natural a las categorías correctas.

6 Conclusiones

Para la empresa, el desarrollo realizado permitió un primer acercamiento al uso de técnicas de Inteligencia Artificial en su portal de ventas online. Los resultados de este proyecto, le permitirán incorporar a futuro nuevas funcionalidades a su sitio. Se podrá, entre otras funcionalidades: a) realizar marketing de manera personalizada y utilizando mejor los recursos disponibles. El recomendador de publicaciones y el perfilado inteligente de usuarios es una herramienta muy potente a la hora de llegar al usuario de forma personalizada y eficaz. b) mejorar la navegación del sitio, en un futuro es posible que la interfaz se adapte al perfil del usuario mejorando plenamente su experiencia en el sitio. c) generar aplicaciones móviles efectivas y que faciliten la adopción por parte de los usuarios.

Para el grupo de investigación participante (conformado por integrantes de dos institutos pertenecientes a la Facultad de Ciencias Exactas de la UNCPBA), resultó en un puntapié fundamental para la transferencia de conocimiento de forma efectiva a la empresa. El grupo se benefició con la posibilidad práctica de materializar algoritmos investigados, y el acceso a información real que permitió validar nuevas técnicas de personalización. Asimismo, el proyecto realizado consistió en el inicio de una relación entre la empresa y el grupo de trabajo que podrá consolidarse a partir de nuevos proyectos que permitan transferir nuevo conocimiento.

Una valiosa lección aprendida de las interacciones y reuniones constantes entre las partes es cómo el proyecto cobra vida y toma direcciones inesperadas. De las reuniones surgió la necesidad de mejorar el tiempo de publicación para incentivar que los usuarios publiquen sus artículos. En este aspecto, se concluyó que elegir la categoría asociada a un anuncio era una traba importante del proceso. Mientras se avanzaba en el desarrollo de los perfiles de usuario, se vislumbró el impacto que podría tener en mejorar el proceso de publicación, contar con una predicción automática de las categorías a partir del título o descripción de los anuncios. Esto llevó a cambiar el rumbo del proyecto durante la última etapa del mismo, para dedicarle más tiempo a este objetivo.

Referencias

1. CACE - “El comercio electrónico creció un 68% y superó los 1,5 billones de pesos en ventas en 2021” - 15 de Marzo de 2021- cace.org.ar - url: https://cace.org.ar/prensa/el-comercio-electronico-crecio-un-68-y-supero-los-15-billones-de- pesos-en-ventas-en-2021/
2. C. Mihai, C. Popa and P. M. Mircea, "Load Profiling For Gas Stations Using Cluster Techniques", IEEE International Power Electronics and Motion Control Conference (PEMC), Varna, pp. 1041-1048, 2016.

3. J. Yang, C. Liu, M. Teng, M. Liao and H. Xiong, "Buyer Targeting Optimization: A Unified Customer Segmentation Perspective", IEEE International Conference on Big Data (Big Data), Washington, DC, pp. 1262- 1271, 2016.
4. A. D. Rachid, A. Abdellah, B. Belaid, L. Rachid, "Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context", International Journal of Electrical and Computer Engineering (IJECE), ISSN: 2088-8708. Vol 8, No 4: August, 2018.
5. Girish S, Ramamurthy B, Senthilnathan T, "Mining the Web Data for Classifying and Predicting Users' Requests", International Journal of Electrical and Computer Engineering (IJECE), ISSN: 2088-8708. Vol 8, No 4, 2018.
6. Yuan, S.-T., & Tsao, Y. W. (2003). A recommendation mechanism for contextualized mobile advertising. Expert Systems with Applications, 24(4), 399–414.
7. Zaim, H., Haddi, A., & Ramdani, M. (2019). A novel approach to dynamic profiling of e-customers considering click stream data and online reviews. International Journal of Electrical and Computer Engineering (IJECE), Vol. 9, p. 602.
8. Molitor, D., Reichhart, P., Spann, M., & Ghose, A. (2019). Measuring the Effectiveness of Location-Based Pull Advertising: A Randomized Field Experiment. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2645281
9. Vallejos, S., and B. Caimmi. 2016. "Geolocalización E Incidentes de Tránsito a Partir Del Análisis de Sentencias Expresadas En Lenguaje Natural Extraídas de Redes Sociales." Edited by A. Soria L. Berdún. Ingeniería de Sistemas, Facultad de Ciencias Exactas, UNCPBA.
10. Caimmi, B, Vallejos S., Berdun L. Soria A., Amandi A, and Campo M. 2016. "Detección de Incidentes de Tránsito En Twitter." In 2016 IEEE Biennial Congress of Argentina (ARGENCON). https://doi.org/10.1109/argencon.2016.7585327.
11. Vallejos S., Caimmi, B, Alonso D., Vallejos S., Berdun L. and Soria A. 2018. "Comparing detection and disclosure of traffic incidents in social networks: an intelligent approach based on Twitter vs. Waze" Inteligencia Artificial, Iberoamerican Journal of Artificial Intelligence (Iberamia) Vol 21 No 61 ISSN 1988-3064
12. D'Cristófaro, M. S., and A. H. Giannoni. 2016. "Asistencia Inteligente En La Planificación Personalizada de La Movilidad Urbana." In Proceedings de EST'16, Concurso de Trabajos Estudiantiles. Directores: A. Soria y L. Berdun.
13. Vallejos, S.; Armentano, M.; Berdun, L.. 2019. "TourWithMe: Recommending peers to visit attractions together" En Proceedings of RecTour, Workshop on Recommender Systems in Tourism. ACM International Conference on Recommender Systems