

Informe:

Detección de *Dark Patterns* por Webscraping


	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Detección de Dark Patterns por Web scraping

Contenido

1. INTRODUCCIÓN	2
2. OBJETIVOS	5
3. ALCANCE DEL ESTUDIO:	5
4. DESCRIPCIÓN PROPÓSITO:	6
5. METODOLOGÍA:	6
6. RESULTADOS:	10
7. CONCLUSIONES:	19

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 3 de 25
	Detección de Dark Patterns por Web scraping		

1. INTRODUCCIÓN

1.1) Antecedentes:

El marketing y las ventas son unidades fundamentales dentro de un negocio¹. El marketing es todo lo que se hace para persuadir al público y las ventas para cerrar el trato.

El marketing es esencial en el modelo de comercio electrónico, ya que digitalmente lo que se busca es imitar las técnicas del mundo off-line y llevarlas al on-line² y así “convertir”, es decir lograr que un consumidor haga lo que se propone una empresa. Existen muchas técnicas para lograr esa conversión, las que nacen del conocimiento del consumidor, sin embargo, existe una zona gris en que ciertas prácticas digitales buscan confundir al consumidor para conseguir dicha conversión, entre esas están los “*Dark Patterns*”.

Los Dark Paterns, del inglés “Patrones Oscuros” se refieren a prácticas digitales que se utilizan para coaccionar, dirigir y engañar a los usuarios. Son utilizados en los sitios web para lograr que el consumidor haga algo que no quería (como: comprar, registrarse, renovar suscripciones, aceptar términos legales, etc) o que si hubiera entendido no lo habría hecho. Este término fue utilizado por primera vez por el Doctor en Ciencias Cognitivas Harry Brignull en el año 2010³. Se dice que es el lado oscuro de la experiencia de usuario, porque los diseñadores utilizan su gran conocimiento sobre el usuario y sus comportamientos para conseguir sus objetivos comerciales, descartando la satisfacción del cliente⁴.

Las estrategias de *Dark Patterns* utilizan técnicas de neuromarketing⁵ y se aprovechan de cómo los usuarios perciben la información, utilizando sesgos cognitivos. El Doctor Brignull indica que la receta de los *Dark Patterns* son la combinación de la psicología aplicada, pruebas A/B y diseño de interfaz de usuario⁶.

Los *Dark Patterns* han ido adquiriendo importancia debido a la masificación del comercio electrónico y la posibilidad de compra con distintos dispositivos. En este sentido, la posibilidad de elaborar patrones oscuros puede ser mediante diseñadores o inteligencia artificial. Según el artículo *Deception&Desing de Harvard Journal of Law & Technology*⁷:

¹ Fuente: Connex. Agencia de Marketing especializada en el sector B2B. Publicada con fecha: 14.04.2022. Recuperada de: <https://blog.connex.es/marketing-y-ventas-claves-para-coordinar-estrategias-y-generar-mas-oportunidades#:~:text=Marketing%20y%20ventas%20son%20dos,parece%20ser%20el%20eterno%20dilema.>

² Fuente: Agencia Marketing Digital. Publicada sin fecha. Recuperada de: <https://www.mdmarketingdigital.com/que-es-el-marketing-digital>

³ Harry Brignull, Dark Patterns: Dirty Tricks Designers Use to Make People Do Stuff, 90 PERCENT OF EVERYTHING (July 8, 2010).


⁴ Fuente: Página Web Estudios Torres Burriel. Artículo: ¿Qué son los Dark Patterns?. Experiencia Usuario. Publicado con fecha: 31.03.2021. Recuperado de: <https://www.torresburriel.com/weblog/2021/03/30/que-son-los-dark-patterns/>

⁵ Neuromarketing es la unión entre la ciencia y las técnicas de marketing. Con base en los estudios relativos a los procesos cerebrales y cómo se desencadenan, el neuromarketing busca medir actividades neurobiológicas cuando hay exposición a marcas, propagandas y otras estrategias de marketing.

⁶ Fuente: Página Web Harry Brignull. Publicada: 06/06/2021. Recuperada de: <https://harrybr.medium.com/bringing-dark-patterns-to-light-d86f24224ebf>

⁷Volume 34, Number 1 Fall 2020

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 4 de 25
	Detección de Dark Patterns por Web scraping		

“El Big Data, seguimiento, aprendizaje automático y otros tipos de inteligencia artificial dan cada vez más forma a las interacciones comerciales con los consumidores. Mediante algoritmos, las empresas emplean herramientas para diseñar publicidad, portales de venta, procesos de devolución y cancelación, precios, productos y servicios propios.

En última instancia, estos algoritmos están programados para optimizar las ganancias. Al mismo tiempo las interfaces digitales poder explorar las características del entorno en línea para manipular y engañar, un fenómeno tan común que el término “Patrones Oscuros” ha sido acuñado para ello.

Aunque los patrones oscuros pueden ser intencionalmente programados, los sistemas de aprendizaje automático actuales pueden aprender por sí mismos a engañar a la gente, incluso cuando los humanos nos los han diseñado para hacerlo”.

La revisión bibliográfica y la experiencia de SERNAC en el tema⁸, indican que la detección de *Dark Patterns* puede realizarse mediante dos técnicas: levantamiento manual o automático (mediante Web Scraping). En el primer caso, un analista experto revisa detalladamente las páginas web de una muestra de proveedores y los categoriza según el tipo de *Dark Pattern*. En el levantamiento automático se construye un algoritmo que detecta frases típicas de cada *Dark Patterns* (traducidos en lenguaje SQL) utilizando Web Scraping. Algunos estudios como el de la Universidad de Princeton⁵ combinan ambos métodos utilizando en primera instancia el método automatizado y luego se realiza una revisión manual que confirma, descarta o re-categoriza la clasificación de *Dark Pattern* que hizo cierto algoritmo.

En el presente estudio de SERNAC en la detección de *Dark Patterns* se utiliza justamente un método combinado mencionado anteriormente.

1.2) Tipos y clasificación de *Dark Patterns*:

Los *Dark Patterns* se clasifican de acuerdo con el efecto que quieren lograr en el consumidor. En ese sentido y según los diversos autores existen entre 15 y 17 tipos.

La siguiente tabla muestra los tipos de *Dark Patterns* que plantea un Estudio de la Universidad de Princeton⁶ y su definición es la que se muestra en la tabla adjunta (Tabla N°:1), en parte serán considerados en el presente estudio, en la medida que puedan ser detectados por texto.

La categoría (primera columna de la tabla) se refiere a la dimensión en que se puede clasificar el *Dark Pattern* y considera entre otros aspectos, características de los patrones y sesgos cognitivos.

Tabla N°1: Clasificación y definición de *Dark Patterns*.

Categoría	Tipo	Descripción
Furtivo/Colarse	Colarse en la cesta	Agregar productos adicionales a los carritos de compras de los usuarios sin su consentimiento.
	Costos escondidos	Revelar cargos previamente no revelados a los usuarios justo antes de realizar una compra.
	Suscripciones escondidas	Cobrar a los usuarios una tarifa recurrente bajo el pretexto de una tarifa única o una prueba gratuita.

⁸ Fuente: Página Web Servicio Nacional del Consumidor (SERNAC). Publicado con fecha: 03.08.2021. Recuperado de: <https://www.sernac.cl/portal/619/w3-article-62983.html>

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Detección de Dark Patterns por Web scraping

Categoría	Tipo	Descripción
Urgencia	Contador de tiempo	Indicar a los usuarios que una oferta o descuento caducará usando un temporizador de cuenta regresiva.
	Mensajes de tiempo límite	Indicar a los usuarios que un trato o venta caducará pronto sin especificar una fecha límite.
Distracción	Confirma Vergüenza (Confirmshaming)	Usar el lenguaje y la emoción (vergüenza) para alejar a los usuarios de hacer una determinada elección.
	Interferencia visual	Usar el estilo y la presentación visual para dirigir usuarios hacia o fuera de ciertas opciones.
	Preguntas capciosas	Usar un lenguaje confuso para orientar a los usuarios en hacer ciertas elecciones.
	Presión de compra	Preseleccionar variaciones más caras de un producto, o presionar al usuario para que acepte las variaciones más caras de un producto y productos relacionados.
Prueba social	Mensajes de actividad	Informar al usuario sobre la actividad en el sitio web (por ejemplo, compras, visualizaciones, visitas).
	Testimonios dudosos	Testimonios en una página de producto cuyo origen no está claro.
Escasez	Mensajes de poco stock	Indicar a los usuarios que cantidades limitadas de un producto están disponibles, aumentando su conveniencia.
	Mensajes de alta demanda	Indicar a los usuarios que un producto tiene una gran demanda y es probable que se agote pronto, lo que aumenta su conveniencia.
Obstrucción	Difícil de cancelar	Hacer que sea fácil para el usuario registrarse en un servicio, pero difícil cancelarlo.
Acciones forzadas	Inscripción forzosa	Obligar a los usuarios a crear cuentas o compartir su información para completar sus tareas.

Fuente: *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*⁶.

1.3) Dark Pattern y asimetría de información:

Los *Dark Patterns* tienen efectos que “En el mejor que los casos molestan y frustran a los usuarios. En el peor de los casos, pueden engañar a los usuarios causando pérdidas financieras, renuncien a gran cantidad de datos personales o inducir a conductas compulsivas y adictivas en adultos y niños”⁹. Los patrones oscuros existen porque son eficaces y son rentables⁵. El uso de *Dark Patterns* se considera una asimetría de información en el mercado, esto se produce cuando una de las partes de una posible transacción maneja más información que la otra de un producto o servicio. En economía se dice lo que provoca una ruptura de la competencia perfecta, conduciendo a un fallo del mercado por ende un resultado económico ineficiente.


En el caso de los *Dark Patterns* quienes manejan más y mejor información son los proveedores, es decir, saben bien cuales son estrategias de marketing y los sesgos de conocimiento que pueden llevar a un consumidor a elegir una alternativa por sobre otra.

Por lo anterior, se considera que los *Dark Patterns* no se encuentran alineados con los derechos que la LPC establece, principalmente, en lo que se refiere a información veraz y oportuna. Lo anterior constituye una asimetría de información en desventaja para el consumidor. La asimetría de información tiene las siguientes consecuencias:

- ☒ Diferencia la percepción de las características del producto.
- ☒ Pone al miembro con más información en una situación privilegiada, sin embargo, conduce a una falla en el mercado, que proporciona un resultado económico ineficiente. Siempre se analiza respecto de la

⁹Fuente: *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*. Universidad de Princeton. arXiv:1907.07032v2 [cs.HC] 20 Sep 2019.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 6 de 25
	Detección de Dark Patterns por Web scraping		

relación proveedor-consumidor, pues las implicancias también son para la relación entre proveedores por la falta de transparencia en el mercado y competencia desigual.

El rol que tiene la Ley del Consumidor es reducir la brecha de conocimiento del consumidor acerca de lo que adquiere a través de las obligaciones de información y de un conjunto equilibrado de derechos y obligaciones que la relación de consumo implica¹⁰. La asimetría de información se equilibra con acciones con diferentes actores: educación para los consumidores, mejorar las buenas prácticas y la autorregulación en los proveedores, mejorar la regulación y/o fiscalización por parte de las instituciones del estado vinculadas.

La protección al consumidor es un instrumento de superación de la desigualdad, del desequilibrio de proveedores y consumidores. Además, sirve para superar la asimetría informativa cuando ello existe.

1.4) Estudios anteriores:

En agosto del 2021 SERNAC realizó un estudio de *Dark Patterns* en el que se revisaron 107 páginas web de empresas con comercio electrónico y se detectó que un 64% (69 empresas) de ellas utiliza al menos un Dark Pattern. El estudio fue de carácter exploratorio con enfoque cuantitativo y cualitativo de temporalidad transversal, es decir, la recolección de datos es en un solo momento del tiempo⁷ y la detección de dichos patrones se hizo de manera manual con la ayuda de funcionarios del mismo servicio que fueron capacitados para dicha actividad. En general, de los tipos de *Dark Pattern* más presentes en las empresas hallados en el estudio, se encuentran las **acciones forzadas** (30 tiendas); **señales de urgencia y escasez** (23 tiendas); y **testimonios dudosos** (18 tiendas)⁸.

Dentro de las empresas que presentaron mayor cantidad de patrones (**4 tipos de Dark Patterns**) están **Travel Security, Froens, Despegar.com, DirecTV, La Polar y Mundo Aromas**⁷.

2. OBJETIVOS

2.1) Objetivo general:


Determinar cuáles son los *Dark Patterns* que pueden ser identificados automáticamente con un algoritmo enfocado en detección de patrones de texto y evaluar su efectividad

2.2) Objetivos específicos:

- Identificar qué *Dark Patterns* pueden ser encontrados utilizando detección de patrones de texto.
- Construir un algoritmo de detección de *Dark Patterns* en texto de páginas web.
- Comparar la manera más efectiva de recoger datos de texto, sea a través fotos de las páginas web o directamente desde el código HTML.

¹⁰ Página Web Diario Digital Ciper Chile. Publicado con fecha: 19.11.2007. Revisado con fecha: 17.05.2021. Recuperado de: <https://www.ciperchile.cl/2007/11/19/derechos-del-consumidor-y-las-brechas-en-el-mercado-del-credito/>

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 7 de 25
	Detección de Dark Patterns por Web scraping		

- Comparar la efectividad que una persona capacitada tiene para detectar *Dark Patterns* contra el algoritmo construido.

3. ALCANCE DEL ESTUDIO:

La investigación realizada es de tipo exploratorio, porque pretende establecer la prevalencia de los *Dark Patterns* en una muestra de páginas webs, con el fin de evidenciar conductas que puedan afectar el conjunto de derechos y obligaciones que la relación de consumo implica para el consumidor y su efecto en una compra informada.

4. DESCRIPCIÓN PROPÓSITO:

El estudio tiene como propósito realizar una herramienta automática que permita generar un diagnóstico de la prevalencia y frecuencia de los *Dark Patterns*, con el fin de establecer posibles infracciones y malas prácticas que afecten la decisión de compra informada del consumidor, con el objeto de proponer correcciones y colaborar en la transparencia del mercado.

5. METODOLOGÍA:

5.1) Resumen metodología:

En la siguiente tabla se indica el resumen de la metodología seguida para este estudio:

Tabla N°2: Resumen metodológico de las distintas muestras.

Categoría	Muestra de empresas	Muestra de Url por Webscraping	Muestra de Url para revisión Manual
Fecha de referencia de los Datos	Inicio: 2.11.2022 Termino: 4.11.2022	Inicio:15.11.2022 Termino: 16.11.2022	Inicio:15.11.2022 Termino: 16.11.2022
Cobertura de establecimientos	Empresas: Revisión de 42 seleccionadas a través de muestreo aleatorio del ranking global de Netrica.	Empresas: 42 revisadas. Disponibilidad de Sitemap: Selección del total de las Url contenidas en el sitemap, para luego definir el tamaño de las Url que serán procesadas por el algoritmo. Sin disponibilidad de Sitemap: Selección de la totalidad de las Url disponibles en el home principal del dominio.	En base al total de Url's obtenidas del proceso por Web Scraping. Se obtiene una muestra aleatoria de los 42 dominios de los establecimientos anteriormente definidos.
Tamaño de la muestra	42 empresas	3.994 Url. Obtenidas de los sitios web de las 42 empresas	360 Url. Obtenidas del total de Url identificadas (3.994 Urls)

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Tipo de muestra	Selección de Empresas: Aleatorio	Selección de Url: Aleatorio	Selección de Url: Aleatorio
Modalidad de levantamiento de información	Se seleccionan las empresas del ranking global de Netrica (año 2022).	Identificación de Url a través de los sitemap o del home de cada sitio web. Obtención completa del body de html de cada url. Screenshot completo de cada url utilizando selenium. Se procesa la imagen obtenida utilizando Tesseract para capturar el texto de su contenido. Uso de expresiones regulares para la identificación de <i>Dark Patterns</i> en el body de html y el texto capturado por Tesseract.	En base al screenshot completo de cada url, se revisa de forma manual el contenido del sitio para detectar los <i>Dark Patterns</i>
Característica del dato	Empresas con páginas Web y venta on-line.	Expresiones regulares para identificar Patrones de texto.	Criterio de analistas expertos capacitados para definir presencia de Dark Pattern en Url
Tipo de análisis	Análisis comparativo / descriptivo	Análisis comparativo / descriptivo	Análisis comparativo / descriptivo
Selección de datos	Selección aleatoria de 42 empresas del ranking global de Netrica.	Selección aleatoria de las url detectadas a través de la búsqueda en sitemap o home de cada dominio web	Selección aleatoria de la muestra de urls detectadas por el Algoritmo.

Fuente: SERNAC, 2022.

5.2) ¿Qué se hizo?

Con el fin de simplificar la comprensión del proceso de detección de *Dark Patterns* por Web scraping, se elaboró este esquema, las etapas se describen de manera sencilla con el objetivo de aclarar el proceso más que plasmar definiciones técnicas.

Imagen N°1: Esquema del proceso de detección de *Dark Patterns* por Webscraping.



Fuente: SERNAC, 2022.

5.3) Definición de las muestras:

El detalle de la definición de las muestras se encuentra en el Anexo N°1.

5.4) *Dark Patterns* detectados por texto mediante Webscraping:

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Si bien la taxonomía de *Dark Patterns* es la que se propone en el punto anterior, sólo algunos de los diferentes *Dark Patterns* pueden ser detectados mediante texto.

Para poder detectar *Dark Patterns* por texto, se realizó un ejercicio de recolección de frases características, en las que se pudieron rescatar 360 frases tipo correspondientes a 12 *Dark Patterns*. Lo anterior, implica que pueden ser extraídas y recolectadas por Web Scraping (reconociendo una frase transcrita a lenguaje SQL).

Respecto a los 6 *Dark Patterns* que no se pueden detectar vía Web Scraping (destacados en rojo en la tabla N°3), estas requieren el análisis integral de otros factores adicionales a frases típicas, como lo son: ejercicio de compra, detección de colores, evaluación de patrones de descripción de productos, lectura sistemática de anuncios completos, entre otros factores.

Tabla N°3: Distribución de *Dark Patterns* por tipo de la matriz de frases tipo.

ID	Tipo de Dark Pattern	Estado Levantamiento Frases	Subtotal Frases
1.	Urgencia	Terminado	96
2.	Escasez	Terminado	41
3.	Notificaciones de actividad	Terminado	17
4.	Roach Motel	Terminado	16
5.	Obstáculos en la comparación de precios	Sin levantar	No aplica
6.	Precios por goteo o costos ocultos	Terminado	11
7.	Colarse en el carrito de compra	Sin levantar	No aplica
8.	Desviar atención o desorientar a los consumidores	Sin levantar	No aplica
9.	Acciones forzadas	Terminado	34
10.	Continuidad forzada o suscripciones ocultas	Sin levantar	No aplica
11.	Presión de compra	Terminado	7
12.	Anuncios disfrazados	Sin levantar	No aplica
13.	Preguntas capciosas	Terminado	7
14.	Testimonios dudosos	Terminado	16
15.	Confirma vergüenza (Confirmshaming)	Terminado	45
16.	Cebo y cambio	Sin levantar	No aplica
17.	Prueba social (Spam de amigos)	Terminado	58
18.	Privacidad de Zuckering	Terminado	10
TOTAL			360

Fuente: SERNAC, 2022.


5.5) Herramienta automatizada:

La herramienta automatizada presentada en este informe busca patrones de texto que determinen o sean una advertencia (denominado como warning de aquí en adelante) de un *Dark Pattern*.

Algunos ejemplos de textos que podrían corresponder a este tipo de interfaces (*Dark Patterns*) podrían ser: ¡Últimas unidades!, ¡Apúrate! y compra ya!, entre otros. Considerando esto y como input principal al algoritmo se construyó una matriz de 360 patrones de texto (frases), correspondientes a 12 tipos distintos de *Dark Patterns*, (Tabla N°3, nombres en negro).

Tomando en cuenta que hay algunas situaciones donde el texto (frase tipo) no basta para determinar de manera concluyente si podría corresponder a un *Dark Pattern*, se clasifican algunos de estos patrones como

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 10 de 25
	Detección de Dark Patterns por Web scraping		

“Warnings” (Advertencias). Los que corresponden a frases que están vinculadas con temas como: cookies (en ocasiones disponibilidad y acceso de datos personales), Cyber Day y Black Friday. En los 2 últimos casos están relacionadas con eventos Cyber, cuyo nombre se configuró para ese fin, aun cuando no en todos los casos se utiliza en fechas Cyber por lo que requiere revisión en detalle, sistemática y caso a caso.

De los 360 patrones de texto (frases) se consideran como Warning’s, 10 corresponden a esta clasificación.

Además, se tomaron decisiones en el proceso de construcción del algoritmo según la disponibilidad de sitemap para encontrar urls dentro de cada dominio. En caso de lograr encontrar el sitemap del dominio se utiliza un muestreo aleatorio simple de las urls recogidas del sitemap. En el caso de que no se encuentre, se recaban todas las urls que se encuentren como links del home principal del dominio.

Otro punto relevante en la definición del algoritmo fueron las tecnologías utilizadas para levantar y procesar los datos. Como lenguaje de programación se utilizó python, por la facilidad que tiene para poder levantar datos y analizar desde páginas web.

Para automatizar un navegador, en concreto Google Chrome, se utilizó Selenium considerando en que tenga la capacidad de ejecutar código. De esta manera se levantó el texto de dos formas:

- Primera: el atributo texto dentro de todo el body de html de cada url levantada, con la desventaja de que el texto de las imágenes no se puede extraer. Segundo, utilizando screenshots de Selenium y Tesseract para transformar fotos a texto.
- Segunda: corresponde a un software entrenado para detectar texto desde fotografías con inteligencia artificial.

De esta manera se logra abordar el problema del primer método de levantamiento. El análisis se desarrolló principalmente con la librería Pandas.

Para terminar, se presenta un diagrama que señala los pasos que toma el algoritmo para recoger los datos y procesar los datos de manera automática.

5.6) Evaluación del algoritmo:

Para poder evaluar el rendimiento del algoritmo de detección de *Dark Patterns*, se plantea una revisión manual desde una sub-muestra de las páginas levantadas de manera automática.

Esta sub-muestra corresponde a 360 páginas tomadas aleatoriamente de las 3.994 páginas revisadas por el algoritmo automatizado. Se dividen en tres analistas capacitados para revisar y clasificar los patrones de texto que se encuentren en las páginas.

De esta manera se medirá la precisión del algoritmo, considerando los tres indicadores más comunes en la clasificación de datos: “Recall” (exhaustividad), precisión (precisión) y “accuracy” (exactitud).

Estos indicadores buscan determinar qué tan bien están identificados los *Dark Patterns*, al comparar contra el levantamiento manual. A continuación, se usa la siguiente nomenclatura: VP corresponde a los verdaderos positivos, FN corresponde a falsos negativos, TN corresponde a falsos negativos y FP corresponde a falsos positivos. El cálculo de los indicadores es el siguiente:

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Tabla N°4: Indicadores de evaluación del algoritmo.

Nombre Indicador	Definición	Formula
Accuracy (exactitud)	Mide que tan preciso es el algoritmo en la clasificación de manera general.	$Accuracy = \frac{VP+VN}{VP+VN+FP+FN}$
Recall (exhaustividad)	Mide que tan bien puede predecir la variable objetivo respecto de todos los casos relevantes a predecir.	$Recall = \frac{VP}{VP+FN}$
Precisión	Mide que tan bien logra predecir el algoritmo considerando los errores de predicción en la variable objetivo del algoritmo.	$Precision = \frac{VP}{VP+FP}$


Fuente: SERNAC, 2022.

6. RESULTADOS:

6.1) Resumen de resultados:

- ✓ **Resultados “Número de Dark Patterns por url método texto html”:** Se puede observar que con una prevalencia de 22,91% de las urls analizadas, poseen entre 1 y 5 *Dark Patterns* según el algoritmo. Algunas incluso alcanzan 5 casos en una misma url, es el caso de las empresas Dafiti y La Polar.
- ✓ **Resultados “Número de Dark Patterns por url método texto foto”:** Se puede observar que en este caso hay una disminución en los *Dark Patterns* en cerca de un 5%, presentando una prevalencia de 17,53% respecto al caso anterior. Al revisar algunos casos se encuentra que el algoritmo no detecta adecuadamente algunas palabras que si son encontradas en el código html (por idioma, límite del diccionario del algoritmo). Cabe mencionar que hay casos en que detecta texto en imágenes que el método anterior no permite encontrar.
- ✓ **Resultados “Número de Dark Patterns por url ambos métodos”:** Podemos observar que en este caso hay un 23,28% de prevalencia, es decir un aumento de 0,37% de urls que presentan *Dark Patterns* respecto al primer método, esto es provocado por los casos en que se logró identificar a través de fotos y no con el código html.
- ✓ **Resultados “Warnings por url método texto html”:** Se observa que se presenta una prevalencia de 33% de las urls que poseen al menos un Warning en texto html.
- ✓ **Resultados “Warnings por url método texto foto”:** En este caso vemos que la prevalencia se acerca al 24%, nuevamente el análisis de foto tiende a equivocarse al traspasar el texto.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 12 de 25
	Detección de Dark Patterns por Web scraping		

- ✓ **Resultados “Warnings por url método texto ambos métodos”:** En este grupo de patrones no presentó ninguna diferencia respecto al uso de texto de html para el análisis. A modo de resumen, la prevalencia es de 23,31% de urls, las que cuentan con al menos un Dark Pattern identificado por la herramienta automática.
- ✓ **Resultados de patrones de texto que identifican Dark Patterns:** Se puede observar que envió gratis (13,13%), % de descuento off (12,56%) y viendo (10,44%) corresponden a la mayor cantidad de *Dark Patterns* encontrados.
- ✓ **Resultados de patrones de texto que identifican Warnings:** Se observa que Cookies (42,60%), Black Friday (23,37%) y Cyber Monday (12,65%) corresponden a la mayor cantidad de Warnings.
- ✓ **Resultados manuales para evaluar el algoritmo de Dark Patterns:** De las url’s revisadas, se encontraron 94 páginas presentaban uno o más *Dark Patterns*, representando el 23,33% del total de la muestra, catalogados principalmente 5 tipos de los 12 buscados. Liderados por los *Dark Patterns* de urgencia (48,93%), preguntas capciosas (25,53%) y testimonios dudosos (12,76%).
- ✓ **Comparación de resultados:** Para evaluar el desempeño del algoritmo se debe definir una sub-muestra de url’s y comparar la detección automatizada con la manual. Las métricas de esta comparación indican que en un análisis considerando las frases *Dark Patterns* y Warnings es más exhaustiva (78,7%) pero menos precisa (51,4%) y exacta (75,0%). Por su parte, al considerar sólo frases *Dark Patterns* el desempeño del algoritmo es más exacto (84,7%) y preciso (76,7%), menos exhaustivo (56,9%) y con mayor rendimiento combinado (Fscore= 67,0%).

6.2) Resultados de detección de *Dark Patterns* de manera automatizada:

A continuación, se presentarán los resultados obtenidos utilizando la herramienta construida para la detección automática de *Dark Patterns*. Del total de empresas se evidenciaron dos casos de empresas que ya no existen y 2 que bloquearon el buscador automatizado. Con esto, se logró levantar datos de 42 empresas distintas y 41 dominios distintos. Lo anterior en un plazo de dos días de manera manual (15 y 16 de noviembre del 2022). La cantidad de urls distintas levantadas corresponde a un total de 3.994.

Se hace hincapié que hay algunas situaciones donde el texto (frase tipo) no basta para determinar de manera concluyente si podría corresponder a un *Dark Pattern*, se clasifican algunos de estos patrones como “Warnings” (Advertencias). Los que corresponden a frases que están vinculadas con temas como: cookies (en ocasiones disponibilidad y acceso de datos personales), Cyber Day y Black Friday. En los 2 últimos casos están relacionadas con eventos Cyber, cuyo nombre se configuró para ese fin, aun cuando no en todos los casos se utiliza en fechas Cyber por lo que requiere revisión en detalle, sistemática y caso a caso.

Los resultados se dividen en cantidad de *Dark Patterns* por url y por patrón de texto. El análisis de patrones por url a su vez se separa en dos partes, primero, un análisis de patrones de texto que identifican *Dark Patterns*, luego para las frases catalogadas como: “Warnings”.

Por lo anterior, los resultados se presentan en el siguiente orden:

6.2.1) Resultados “Número de *Dark Patterns* por url método texto html”.

6.2.2) Resultados “Número de *Dark Patterns* por url método texto foto”.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

- 6.2.3) Resultados “Número de *Dark Patterns* por url ambos métodos”.
- 6.2.4) Resultados “Warnings por url método texto html”.
- 6.2.5) Resultados “Warnings por url método texto foto”.
- 6.2.6) Resultados “Warnings por url método texto ambos métodos”.
- 6.2.7) Resultados de patrones de texto que identifican *Dark Patterns*.
- 6.2.8) Resultados de patrones de texto que identifican Warnings.
- 6.3) Resultados manuales para evaluar el algoritmo de *Dark Patterns*.
- 6.4) Comparación de resultados.

Los resultados en detalle se muestran a continuación:

6.2.1) Resultados “Número de *Dark Patterns* por url método texto html”:

Se puede observar que con una prevalencia de 22,91% de las urls analizadas poseen entre 1 y 5 *Dark Patterns*, según el algoritmo. Algunas incluso alcanzan 5 casos en una misma url, es el caso de las empresas Dafiti y La Polar.

Tabla N°5: Número de *Dark Patterns* por url método texto html.

Nº <i>Dark Patterns</i> texto html	Cuenta de url	% del total
0	3.079	77,09%
1	566	14,17%
2	171	4,28%
3	64	1,60%
4	112	2,80%
5	2	0,05%
Total general	3.994	100%

Fuente: SERNAC, 2022.

6.2.2) Resultados “Número de *Dark Patterns* por url método texto foto”:

Se puede observar que en este caso hay una disminución en los *Dark Patterns* en cerca de un 5%, presentando una prevalencia de 17,53% respecto al caso anterior. Al revisar algunos casos se encuentra que el algoritmo no detecta adecuadamente algunas palabras que si son encontradas en el código html (por idioma, límite del diccionario del algoritmo). Cabe mencionar que hay casos en que detecta texto en imágenes que el método anterior no permite encontrar.

Tabla N°6: Número de *Dark Patterns* por url método texto foto.

Nº <i>Dark Patterns</i> texto foto	Cuenta de url	% del total
0	3.294	82,47%
1	430	10,77%
2	126	3,15%
3	96	2,40%
4	46	1,15%
5	2	0,05%
Total general	3.994	100%

Fuente: SERNAC, 2022.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

6.2.3) Resultados “Número de *Dark Patterns* por url ambos métodos”:

Podemos observar que en este caso hay un 23,28% de prevalencia, es decir un aumento de 0,37% de urls que presentan *Dark Patterns* respecto al primer método, esto es provocado por los casos en que se logró identificar a través de fotos y no con el código html.

Tabla N°7: Número de *Dark Patterns* por url ambos métodos.

Nº <i>Dark Patterns</i> html o foto	Cuenta de url	% del total
0	3.064	76,72%
1	576	14,42%
2	163	4,08%
3	69	1,73%
4	119	2,98%
5	3	0,08%
Total general	3.994	100%

Fuente: SERNAC, 2022.

6.2.4) Resultados “Warnings por url método texto html”:

Se observa que se presenta una prevalencia de 33% de las urls que poseen al menos un Warning en texto html.

Tabla N°8: Warnings por url método texto html.

Nº Warning Dark Pattern texto html	Cuenta de url	% del total
0	2.702	67,65%
1	931	23,31%
2	291	7,29%
3	70	1,75%
Total general	3.994	100%

Fuente: SERNAC, 2022.

6.2.5) Resultados “Warnings por url método texto foto”:


En este caso vemos que la prevalencia se acerca al 24%, nuevamente el análisis de foto tiende a equivocarse al traspasar el texto.

Tabla N°9: Warnings por url método texto foto.

Nº Dark Pattern texto foto	Cuenta de url	% del total
0	3.050	76,36%
1	887	22,21%
2	56	1,40%
3	1	0,03%
Total general	3.994	100%

Fuente: SERNAC, 2022.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 15 de 25
	Detección de Dark Patterns por Web scraping		

6.2.6) Resultados “Warnings por url método texto ambos métodos”:

En este grupo de patrones no presentó ninguna diferencia respecto al uso de texto de html para el análisis. A modo de resumen, la prevalencia es de 23,31% de urls, las que cuentan con al menos un Dark Pattern identificado por la herramienta automática.

Tabla N°10: Warnings por url método texto ambos métodos.

Nº Warning Dark Pattern html o foto	Cuenta de url	% del total
0	2,702	67,65%
1	931	23,31%
2	291	7,29%
3	70	1,75%
Total general	3.994	100%

Fuente: SERNAC, 2022.

Además, cerca de un 32,35% presentan entre 1 y 3 Warnings. Por último, los dos métodos de identificación de texto se complementan, pero, en general, las fotos solo funcionan mejor en un caso acotado de casos respecto del total.

6.2.7) Resultados de patrones de texto que identifican Dark Patterns:

Se puede observar que envió gratis (13,13%), % de descuento off (12,56%) y viendo (10,44%) corresponden a la mayor cantidad de Dark Patterns encontrados. Cabe resaltar que algunos de estos patrones podrían estar identificando erróneamente, pero esto se revisará más adelante en la comparación con el método manual de identificación. Además, cabe resaltar que se encontraron la presencia de 38 frases de Dark Patterns versus las 360 frases tipo que se utilizaron.

Tabla N°11: Cantidad de Dark Patterns encontradas por patrón de texto.

Frase Dark Pattern	Patrón de texto (frase característica transformada a lenguaje SQL)	Dark Patterns en texto Html		Dark Patterns en foto		Dark Patterns en Html o foto	
		Nº	%	Nº	%	Nº	%
Envío gratis	(^ ?env[i f]o gratis(\n? ?) env[i f]o gratis\$?env[i f]o gratis(\n)?)	210	13,48%	106	9,11%	210	13,13%
[número] % off	(^ ?\d+% off(\n? ?) \d+% off\$?\d+% off(\n)?)	197	12,64%	185	15,89%	201	12,56%
Viendo	(^ ?viendo(\n? ?) viendo\$?viendo(\n)?)	167	10,72%	107	9,19%	167	10,44%
Descuento exclusivo	(^ ?descuento exclusivo(\n? ?) descuento exclusivo\$?descuento exclusivo(\n)?)	141	9,05%	139	11,94%	141	8,81%
Sólo por hoy	(^ ?s[ó o]lo por hoy(\n? ?) s[ó o]lo por hoy\$?s[ó o]lo por hoy(\n)?)	110	7,06%	110	9,45%	110	6,88%

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Detección de Dark Patterns por Web scraping

Frase Dark Pattern	Patrón de texto (frase característica transformada a lenguaje SQL)	Dark Patterns en texto Html		Dark Patterns en foto		Dark Patterns en Html o foto	
		N°	%	N°	%	N°	%
¡Mejor precio!	(^ ?mejor precio(\n? ?) mejor precio\$?mejor precio(\n)?)	109	7,00%	65	5,58%	109	6,81%
Excelente servicio	(^ ?excelente servicio(\n? ?) excelente servicio\$?excelente servicio(\n)?)	106	6,80%	109	9,36%	109	6,81%
[número] vendidos	(^ ?\d+ vendidos(\n? ?) \d+ vendidos\$?\d+ vendidos(\n)?)	94	6,03%	41	3,52%	94	5,88%
Por tiempo limitado	(^ ?por tiempo limitado(\n? ?) por tiempo limitado\$?por tiempo limitado(\n)?)	73	4,69%	61	5,24%	73	4,56%
Últimas unidades	(^ ?[ú u]ltimas unidades(\n? ?) [ú u]ltimas unidades\$?[ú u]ltimas unidades(\n)?)	65	4,17%	15	1,29%	67	4,19%
Stock limitado	(^ ?stock limitado(\n? ?) stock limitado\$?stock limitado(\n)?)	46	2,95%	15	1,29%	46	2,88%
Quedan [numero] unidades	(^ ?(i)?quedan \d+ unidades(!)?(\n? ?) (i)?quedan \d+ unidades(!)?\$?(i)?quedan \d+ unidades(!)?(\n)?)	37	2,37%	37	3,18%	37	2,31%
Compró	(^ ?compr[o ó](\n? ?) compr[o ó]\$?compr[o ó](\n)? ?compr[o ó](\.\.)?\$(\n)?)	25	1,60%	10	0,86%	25	1,56%
Excelente calidad	(^ ?excelente calidad(\n? ?) excelente calidad\$?excelente calidad(\n)?)	25	1,60%	27	2,32%	31	1,94%
Es perfecto	(^ ?es perfecto(\n? ?) es perfecto\$?es perfecto(\n)?)	22	1,41%	14	1,20%	23	1,44%
Excelente producto	(^ ?excelente producto(\n? ?) excelente producto\$?excelente producto(\n)?)	14	0,90%	31	2,66%	33	2,06%
Ofertas cyber	(^ ?ofertas cyber(\n? ?) ofertas cyber\$?ofertas cyber(\n)?)	12	0,77%	10	0,86%	12	0,75%
Muy buena calidad	(^ ?muy buena calidad(\n? ?) muy buena calidad\$?muy buena calidad(\n)?)	11	0,71%	16	1,37%	17	1,06%
Hasta agotar stock	(^ ?(i)?hasta agotar stock(!)?(\n? ?) (i)?hasta agotar stock(!)?\$?(i)?hasta agotar stock(!)?(\n)?)	10	0,64%	2	0,17%	10	0,63%
Edición limitada	(^ ?(i)?edici[o ó]n i jmitada(!)?(\n? ?) (i)?edici[o ó]n i jmitada(!)?\$?(i)?edici[o ó]n i jmitada(!)?(\n)?)	9	0,58%	4	0,34%	9	0,56%
Interesado	(^ ?interesado(\n? ?) interesado\$?interesado(\n)?)	9	0,58%	6	0,52%	9	0,56%
Nuestros clientes dicen	(^ ?nuestros clientes dicen(\n? ?) nuestros clientes dicen\$?nuestros clientes dicen(\n)?)	8	0,51%	8	0,69%	8	0,50%
Oferta especial	(^ ?oferta especial(\n? ?) oferta especial\$?oferta especial(\n)?)	8	0,51%	8	0,69%	8	0,50%
Viendo esto	(^ ?viendo esto(\n? ?) viendo esto\$?viendo esto(\n)?)	8	0,51%	3	0,26%	8	0,50%
Fabuloso	(^ ?fabuloso(\n? ?) fabuloso\$?fabuloso(\n)?)	7	0,45%	6	0,52%	7	0,44%

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Detección de Dark Patterns por Web scraping

Frase Dark Pattern	Patrón de texto (frase característica transformada a lenguaje SQL)	Dark Patterns en texto Html		Dark Patterns en foto		Dark Patterns en Html o foto	
		N°	%	N°	%	N°	%
Hasta un [número] % de descuento	(^?hasta un \d+ ?% de descuento(\n? ?) hasta un \d+ ?% de descuento\$?hasta un \d+ ?% de descuento(\n?)?)	7	0,45%	5	0,43%	7	0,44%
Compra ahora	(^?compra ahora(\n? ?) compra ahora\$?compra ahora(\n?)?)	5	0,32%	5	0,43%	5	0,31%
Mega oferta	(^?mega oferta(\n? ?) mega oferta\$?mega oferta(\n?)?)	5	0,32%	5	0,43%	5	0,31%
Solo quedan [Número]	(^ ?(i)?s[o ó]lo quedan \d+(!)?(\n? ?) (i)?s[o ó]lo quedan \d+(!)?\$?(i)?s[o ó]lo quedan \d+(!)?(\n?)?)	3	0,19%	3	0,26%	3	0,19%
Oferta exclusiva	(^ ?oferta exclusiva(\n? ?) oferta exclusiva\$?oferta exclusiva(\n?)?)	3	0,19%	1	0,09%	3	0,19%
Alguien en	(^ ?alguien en(\n? ?) alguien en\$?alguien en(\n?)?)	2	0,13%	2	0,17%	2	0,13%
Alta demanda	(^?alta demanda(\n? ?) alta demanda\$?alta demanda(\n?)?)	2	0,13%	0	0,00%	2	0,13%
Envío rápido	(^?env[í i]o r[a á]pido(\n? ?) env[í i]o r[a á]pido\$?env[í i]o r[a á]pido(\n?)?)	2	0,13%	3	0,26%	3	0,19%
Precio oferta	(^?precio oferta(\n? ?) precio oferta\$?precio oferta(\n?)?)	2	0,13%	2	0,17%	2	0,13%
¡No te lo pierdas!	(^?(i)?no te lo pierdas(!)?(\n? ?) (i)?no te lo pierdas(!)?\$?(i)?no te lo pierdas(!)?(\n?)?)	1	0,06%	0	0,00%	1	0,06%
Happy hour	(^?happy hour(\n? ?) happy hour\$?happy hour(\n?)?)	1	0,06%	1	0,09%	1	0,06%
Reserva hoy	(^?reserva hoy(\n? ?) reserva hoy\$?reserva hoy(\n?)?)	1	0,06%	1	0,09%	1	0,06%
Sí, lo quiero	(^?s[i í](,)? lo quiero(\n? ?) s[i í](,)? lo quiero\$?s[i í](,)? lo quiero(\n?)?)	1	0,06%	1	0,09%	1	0,06%
Total general	38	1.558	100%	1.164	100%	1.600	100%

Fuente: SERNAC, 2022.

6.2.8) Resultados de patrones de texto que identifican Warnings:

Se observa que Cookies (42,60%), Black Friday (23,37%) y Cyber Monday (12,65%) corresponden a la mayor cantidad de Warnings. Considerando la fecha en que se realizó el levantamiento hace sentido la marca de Black Friday, pero no así Cyber Monday, por lo cual en este caso se requeriría re-evaluarlas. En este caso el número de frases Warnings encontrados es 7.

Tabla N°12: Cantidad de Warnings por Patrón de texto.

Frase Dark Pattern	Patrón de texto (frase característica transformada a lenguaje SQL)	Warning Dark Patterns en texto html		Warning Dark Patterns en texto foto		Warning Dark Patterns en html o foto	
		N°	%	N°	%	N°	%
Cookies	(^ ?cookies(\n? ?) cookies\$?cookies(\n?)?)	734	42,60%	426	42,51%	734	42,60%

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Detección de Dark Patterns por Web scraping

Frase Dark Pattern	Patrón de texto (frase característica transformada a lenguaje SQL)	Warning Dark Patterns en texto html		Warning Dark Patterns en texto foto		Warning Dark Patterns en html o foto	
		N°	%	N°	%	N°	%
Black friday	(^ ?black friday(\n? ?) black friday\$?black friday(\n)?)	401	23,27 %	263	26,25 %	401	23,27%
Cyber monday	(^ ?cyber monday(\n? ?) cyber monday\$?cyber monday(\n)?)	218	12,65 %	133	13,27 %	218	12,65%
Descuento exclusivo para ti	(^ ?descuento exclusivo para ti(\n? ?) descuento exclusivo para ti\$?descuento exclusivo para ti(\n)?)	138	8,01%	129	12,87 %	138	8,01%
Éste sitio usa cookies	(^ ?[e é]ste sitio usa cookies(\n? ?) [e é]ste sitio usa cookies\$?[e é]ste sitio usa cookies(\n)?)	124	7,20%	44	4,39%	124	7,20%
Utilizamos cookies	(^ ?utilizamos cookies(\n? ?) utilizamos cookies\$?utilizamos cookies(\n)?)	107	6,21%	6	0,60%	107	6,21%
Utilizamos cookies propias	(^ ?utilizamos cookies propias(\n? ?) utilizamos cookies propias\$?utilizamos cookies propias(\n)?)	1	0,06%	1	0,10%	1	0,06%
Total general	7	1.723	100%	1002	100%	1.723	100%

Fuente: SERNAC, 2022.

6.3) Resultados manuales para evaluar el algoritmo de Dark Patterns:

Para evaluar el algoritmo utilizado en la clasificación automática de *Dark Patterns*, se determinó una submuestra de 360 Url's que fueron revisadas por analistas expertos, que evaluaron manual y sistemáticamente cada uno de los casos y decidieron confirmar, descartar o re-categorizar la clasificación de Dark Pattern que hizo tal algoritmo.

Los resultados fueron los siguientes:

Tabla N°13: Matriz de confusión del algoritmo.

Nº Url con DP Evaluación Manual	Cuenta de url	% del total
0	266	76,67%
1	94	23,33%
Total general	360	100%

Fuente: SERNAC, 2022.

De las url's revisadas, se encontraron 94 páginas presentaban uno o más *Dark Patterns*, representando el 23,33% del total de la muestra, catalogados principalmente 5 tipos de los 12 buscados. Liderados por los *Dark Patterns* de urgencia (48,93%), preguntas capciosas (25,53%) y testimonios dudosos (12,76%).

Tabla N°14: Distribución de frecuencia por tipo de Dark Patterns (sin Warnings).

Tipo de Dark Pattern	Número	% del Total
Urgencia	46	48,93%
Preguntas capciosas	24	25,53%

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Testimonios dudosos	12	12,76%
Escasez	11	11,07%
Inscripción forzosa	1	1,06%
Total	94	100%

Fuente: SERNAC, 2022.

Con este insumo, se comparan los resultados del algoritmo con detección automática versus la revisión manual.

6.4) Comparación de resultados:

Tomando en cuenta que hay algunas situaciones donde el texto no basta para determinar si podría corresponder a un Dark Pattern, se clasifican algunos de estos patrones como “Warnings”, se considera que 10 frases corresponden a esta clasificación, en general están referidos a temas de eventos Cyber y Cookies.

La comparación de la detección automatizada del algoritmo versus la manual demuestra que la coincidencia alcanza un 75,0%. Esto quiere decir que en ese mismo valor la categorización automática y la manual realizada por expertos es la misma. La evaluación de un algoritmo tiene múltiples capas y las coincidencias son la primera.

Tabla N°15: Distribución de coincidencias en submuestra “con Warnings”.

Categoría	N° Url	% Url
No Coincide	90	25,0%
Coincide	270	75,0%
Total general	360	100%

Fuente: SERNAC, 2022.

Aun cuando las coincidencias son altas para evaluar adecuadamente el desempeño del algoritmo se requiere plantear la matriz de confusión (tabla N°15). La matriz evidencia la precisión de un algoritmo para clasificar un conjunto de datos, en cada columna se representa las predicciones de cada clase (1 predice que será positivo y 0 negativo) y la fila representa las instancias en la clase real. Es decir, indica qué tipo de aciertos y errores tiene el modelo.

La matriz de confusión constituye la segunda capa de evaluación de desempeño de un algoritmo. En este caso podemos decir que el modelo tiene más aciertos (74) que errores (70).

Tabla N°16: Matriz de confusión del algoritmo “con warnings”.

	categoría	Algoritmo		Total general
		0	1	
Manual	0	196	70	266
Manual	1	20	74	94
	Total general	216	144	360

Fuente: SERNAC, 2022.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Como tercera capa y para evaluar el desempeño del algoritmo, se debe estimar cuantos datos clasifica, si hizo las predicciones correctas y con qué calidad, se deben calcular métricas que son fórmulas que miden cada uno de esos aspectos (tabla N°17).

Tabla N°17: Métricas de evaluación de desempeño del algoritmo.

Métrica	Valor	Significado
Exactitud	75,0%	Mide los aciertos del algoritmo (clasificador) y la cantidad total de datos (%).
Precisión	51,4%	Mide la calidad del algoritmo en tareas de clasificación (%).
Exhaustividad	78,7%	Cantidad de datos del algoritmo que es capaz de clasificar (%).
FScore	62,0%	Combina las medidas de precisión y exhaustividad en un solo valor, rendimiento combinado (%).

Fuente: SERNAC, 2022.

Al no considerar a los Dark Pattern encontrados por el algoritmo y clasificados como “Warnings” (cookies, cyberday, blackmonday), se obtiene el resultado de la tabla N°18. Se destaca que hay una mayor coincidencia entre los resultados.

Tabla N°18: Distribución de coincidencias en submuestra “sin Warnings”.

Categoría	N Url	%
No Coincide	55	15,3%
Coincide	305	84,7%
Total general	360	100%

Fuente: SERNAC, 2022.

Para visualizar el desempeño del algoritmo adecuadamente se debe utilizar nuevamente una matriz de confusión (Tabla N°19):

Tabla N°19: Matriz de confusión del algoritmo “sin Warnings”.

		Algoritmo		Total general
		0	1	
Manual	0	196	70	266
	1	20	74	94
Total general		216	144	360


Fuente: SERNAC, 2022.

Tabla N°20: Métricas de evaluación de desempeño del algoritmo “sin Warnings”.

Métrica	Valor	Significado
Exactitud	84,7%	Mide los aciertos del algoritmo (clasificador) y la cantidad total de datos (%).
Precisión	76,7%	Mide la calidad del algoritmo en tareas de clasificación (%).
Exhaustividad	56,9%	Cantidad de datos del algoritmo que es capaz de clasificar (%).
FScore	67,0%	Combina las medidas de precisión y exhaustividad en un solo valor, rendimiento combinado (%).

Fuente: SERNAC, 2022.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 21 de 25
	Detección de Dark Patterns por Web scraping		

Al contrastar los resultados obtenidos con el algoritmo (con y sin Warnings), las diferencias observadas permiten ordenar en dos categorías las urls, las que coinciden y que no. Al hacer una revisión de las que no coincidieron, se detectaron los siguientes sesgos algorítmicos:

- Falta de inclusión de frases tipo en el Regex: en la matriz del algoritmo, no estaban considerando ciertas frases tipo de *Dark Patterns* que fueron detectados de manera manual.
- Regex no definidos correctamente: en la matriz de frases tipo de *Dark Patterns* se incluyó la palabra “viendo”, lo que el algoritmo consideraba también “viviendo”. Para subsanar este punto se requiere ajustar la matriz.
- Regex correctos pero que no aplicaban como *Dark Patterns*: en a páginas de compra del supermercado Jumbo aparece, “estas viendo esto”, lo que implicaba que el algoritmo buscando que es “viendo” lo considerara como un *Dark Patterns*, constituyendo un caso de falso positivo.
- *Dark Patterns* que estaban en formato de texto: el uso de Selenium no es del todo confiable, por la inconsistencia del cambio de idioma y/o operación de la herramienta, que puede descartar algunos *Dark Patterns* relevantes, lo que implica una subestimación en la detección.

7. CONCLUSIONES:

✓ **Construir un algoritmo de detección de Dark Patterns en texto de páginas web:**

Los *Dark Patterns* pueden presentarse en varios lugares dentro de un sitio web. Dependen en gran medida de la manipulación de la interfaz, cambiar la jerarquía de los elementos de la interfaz o priorizar ciertas opciones usando colores. Sin embargo, a menudo se presentan muchos *Dark Patterns* en las principales rutas de interacción de los usuarios de los servicios en línea⁷.


Estos *Dark Patterns* comparten rasgos comunes como el texto que muestran⁷. De ahí nace la necesidad de construir una matriz de frases tipo. Por lo que la técnica implementada en esta detección nace de la identificación de frases textuales presentes en cada URL revisada, luego se clasifican por tipo de *Dark Patterns*, a lo que seguido se evalúa una sub-muestra para que un analista experto las revise.

Los resultados de la detección de *Dark Patterns* de manera automatizada indican que el algoritmo construido si funciona, dado que en una muestra de 3.994 url’s se detectaron *Dark Patterns* con una prevalencia 40,0% (1.600 casos) considerando frases de *Dark Patterns*/Warnings y de 43,1% (1.723 casos) sólo considerando frases *Dark Patterns*.

Respecto de las frases Warnings (Advertencias), se observó que: Cookies (42,60%), Black Friday (23,37%) y Cyber Monday (12,65%) corresponden a la mayor presencia. Considerando la fecha en que se realizó el levantamiento hace sentido la marca de Black Friday, pero no así Cyber Monday (7), por lo cual en este caso se requeriría re-evaluarlas de manera específica y particular. Queda en discusión el hecho de que este tipo de frases se extienda en el periodo que se usan más porque representa un posible atractivo publicitario que una fecha particular representativa del nombre “Cyber”.

Aun cuando el funcionamiento del algoritmo detecta una porción de las frases tipo definidas para la detección de *Dark Patterns*, se identificaron algunos sesgos algorítmicos reflejados en el regex: como la detección de

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

	Versión: 0.7	Estado: No Vigente	Página 22 de 25
	Detección de Dark Patterns por Web scraping		

frases que no estaban en la matriz de frases tipo de *Dark Patterns*, definición incorrecta de la frase “viendo”, traducción incorrecta de frases y por último dificultades en la operación de Selenium. Por lo que para elevar aun más el desempeño del algoritmo se requiere ampliar la matriz de frases tipo y modificar la identificación de frases Warnings (Advertencias).

✓ **Determinar cuáles son los Dark Patterns que pueden ser identificados automáticamente con un algoritmo enfocado en detección de patrones de texto y evaluar su efectividad:**

La matriz de frases tipo de *Dark Patterns* incluyó 360 alternativas, en la detección automatizada se identificó la presencia de 38 frases, pertenecientes a 5 tipos de *Dark Patterns*.

Considerando un análisis de frecuencia se tiene que los *Dark Patterns* más usuales (sin considerar frases Warnings) son: Señales de urgencia (891 casos), señales de escasez (266 casos), testimonios dudosos (231 casos), prueba social (211 casos) y confirmshaming (1 caso).

Se puede observar que con una prevalencia de 22,91% de las urls analizadas poseen entre 1 y 5 *Dark Patterns*, según la herramienta. En 2 casos tienen 5 tipos de *Dark Patterns* en una misma url (Dafiti y La Polar).

✓ **Identificar qué Dark Patterns pueden ser encontrados utilizando detección de patrones de texto:**

Para poder detectar *Dark Patterns* por texto, se realizó un ejercicio de recolección de frases características, en las que se pudieron rescatar 360 frases tipo correspondientes a 12 *Dark Patterns*. Esta recolección se hizo revisando aproximadamente 500 páginas.

La recolección de frases características implicaba que podrían ser extraídas y recolectadas por Webscraping (reconociendo una frase transcrita a lenguaje SQL), sin considerar adicionales acciones como: simulación de compra, detección de colores, evaluación de patrones de descripción de productos, lectura sistemática de anuncios completos, entre otros factores.

De la matriz de frases tipo de *Dark Patterns* (360) se logró detectar la presencia de 38 de ellas. En lo específico, las frases de *Dark Patterns* con mayor frecuencia fueron: envío gratis (13,13%), % de descuento off (12,56%) y viendo (10,44%).

Tal presencia de *Dark Patterns* y Warnings demuestra que la publicidad en el comercio electrónico es muy dinámica, estacional, de corta exposición y modificada por fechas coyunturales. Lo anterior, en virtud de que entre la construcción de la matriz de frases tipo y la aplicación del algoritmo transcurrió un corto periodo.

✓ **Comparar la manera más efectiva de recoger datos de texto, sea a través fotos de las páginas web o directamente desde el código HTML:**

La prevalencia de *Dark Patterns* en la detección de texto HTML es de 22,91%, la de texto en fotografía de 17,53% y la que combina ambos métodos es de 23,28%.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

Se evidenció que la detección de *Dark Patterns* de texto en fotografía presenta una limitación, ya que no identifica algunas palabras que si corresponden a *Dark Patterns* por temas de idioma y límites del diccionario del algoritmo.

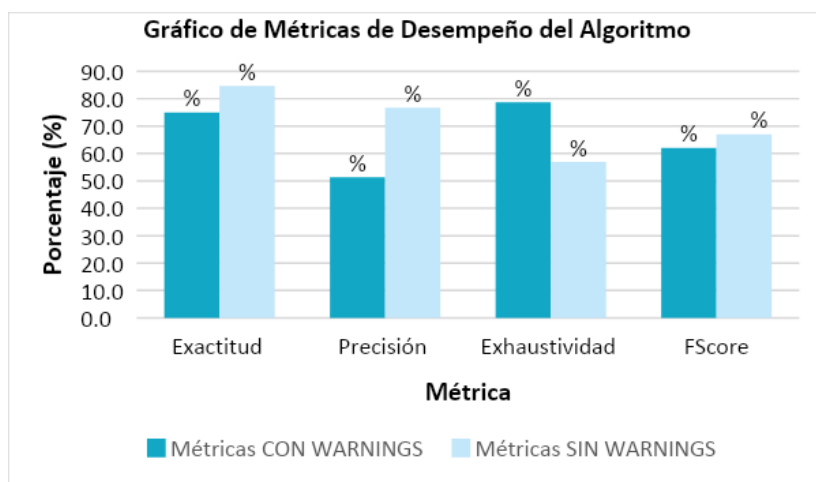
✓ **Comparar la efectividad que una persona capacitada tiene para detectar *Dark Patterns* contra el algoritmo construido:**

La detección manual de *Dark Patterns* realizada por un analista capacitado evidencia un 23,33% (análisis manual) de casos positivos. Por otra parte, al comparar los resultados de la detección manual versus la automatizada se tiene un 84,7% de coincidencias, lo que representa un alto ajuste entre ambos tipos de detecciones.

Las principales desviaciones se dieron por límites de la matriz de frases tipo, la construcción de las frases tipo y en una menor medida la diferencia de criterios de los analistas.

Para evaluar el desempeño del algoritmo se debe definir una sub-muestra de url's y comparar la detección automatizada con la manual. Las métricas de esta comparación indican que en un análisis considerando las frases *Dark Patterns* y *Warnings* es más exhaustiva (78,7%) pero menos precisa (51,4%) y exacta (75,0%). Por su parte, al considerar sólo frases *Dark Patterns* el desempeño del algoritmo es más exacto (84,7%) y preciso (76,7%), menos exhaustivo (56,9%) y con mayor rendimiento combinado (Fscore= 67,0%).

Lo anterior evidencia que la exhaustividad (Cantidad de datos del algoritmo que es capaz de clasificar) del algoritmo está marcada por la presencia de frases *Warnings* que distorsionan y subestiman las métricas de desempeño.




Fuente: SERNAC, 2022.

✓ **Con este estudio se pretende generar los siguientes aportes:**

- Se creó un set de datos que permitió revisar la prevalencia de los *Dark Patterns* en 3.994 sitios webs.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 24 de 25
	Detección de Dark Patterns por Web scraping		


- Contribuir con técnicas de medición automatizadas que permiten a los analistas expertos revisar *Dark Patterns* existentes o detectar nuevos.
- Elevar los mecanismos y magnitudes de monitoreos de publicidad, en consecuencia, generar insumos para una fiscalización masiva.
- Se dispone de una herramienta automatizadas para ayudar a concientizar sobre los *Dark Patterns* a investigadores e instancias reguladoras.
- Se desarrolló una herramienta de cara al consumidor que ayuda la detección de *Dark Patterns* y por ende poder combatirlos.
- Se releva la taxonomía de *Dark Patterns* utilizada en la investigación de la Universidad de Princeton⁶ que proporciona una terminología adecuada para caracterizar el funcionamiento de cada Dark Pattern. Lo que puede ayudar a comprender y comparar la influencia secundaria y los efectos perjudiciales para el consumidor.

8. HALLAZGOS:

Pensando en la utilización de esta herramienta se dan ciertas recomendaciones para la implementación del algoritmo a nivel institucional para fiscalización o estudios posteriores. A continuación, los puntos a considerar:

- a) Dependencias de la herramienta: El algoritmo utiliza librerías de Python (Pandas, Requests, Selenium, etc), Chromedriver y Tesseract OCR. Es necesario estar seguro de que se cuenta con esas herramientas en la máquina en que se utilice.
- b) Ejecución del algoritmo: Idealmente se podría mantener agendado automáticamente en una máquina o un contenedor en la infraestructura del servicio.
- c) Gestión del producto: La herramienta deberá quedar a cargo de un equipo con la capacidad de ejecutar y difundir la información que se produzca desde esta herramienta.
- d) Mejoras al algoritmo: Existen una serie de mejoras ya comentadas que deberían ser abordadas.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 25 de 25
	Detección de Dark Patterns por Web scraping		

ANEXO N°1: DEFINICIÓN DE LAS MUESTRAS

Muestra de Links

Muestreo Aleatorio Simple: Para nuestro caso una vez calculado el número de la muestra, se procede a seleccionar las unidades muestrales desde la población de manera tal que cada n tenga la misma probabilidad de ser seleccionado en la muestra, considerando además que es un muestreo sin reposición, lo que significa que se seleccionara unidad a unidad de observación desde la población sin reponer nuevamente la unidad seleccionada previamente, eliminando el sesgo de que existan repeticiones en la muestra y que las unidades estadísticas sean únicas en la muestra.

El mecanismo de selección en este tipo de método de muestreo es aplicando funciones de números aleatorios, es decir que se escogen las unidades estadísticas, según una asignación equivalente a un número aleatorio por unidad estadísticas cuyas ventajas son las siguientes:

- Sencillo y fácil comprensión
- Cálculo rápido de medias y varianzas
- Existen paquetes informáticos para analizar los datos

En tanto a sus desventajas es que se requiera un listado completo de toda la población o marco muestral y si trabajamos con muestras pequeñas es posible que no representen a toda la población adecuadamente, que en nuestro caso no se tienen estas desventajas dado que no se trabajará con una muestra pequeña y que el marco muestral está definido más adelante.


Muestra de Empresas

Para seleccionar las empresas se utilizará un muestreo no probabilístico, denominado **Muestreo por Juicio**, en donde las unidades estadísticas se eligen sobre la base del conocimiento y el juicio del investigador, dado que el conocimiento del investigador es nuestro caso es el estudio del año 2021 de *Dark Patterns*, hay posibilidades de que los resultados obtenidos sean altamente precisos con un mínimo margen de error, debido a que se escogieron a las empresas de dicho estudio que mayormente presentan esta característica en sus páginas.

Características importantes de este muestreo:

- La muestra es escogida por un experto, de acuerdo con su criterio, la idea es seleccionar las unidades más representativas.
- Se emplea cuando el tamaño de muestra es pequeño.
- En este enfoque de muestreo, la experiencia de los investigadores es importante y no existen otras barreras debido a las cuales la selección de una muestra resulta extremadamente conveniente

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 26 de 25
	Detección de Dark Patterns por Web scraping		

5.4) Diseño muestral:

Diseño muestral para Links

Como se trata de un estudio cuya variable principal es de tipo cualitativo, específicamente del tipo dicotómico (existe ó no existe *Dark Patterns*), dado que el objetivo del estudio es determinar cuales son los *Dark Patterns* que pueden ser determinados a partir de un algoritmo que es desarrollado por la UMM, es que se utilizará el diseño de muestreo aleatorio sistemático, dado que se tiene un orden en la extracción de los links y este tipo de muestreo no es tan necesario mantener un orden respecto de la población, desde internet, por intermedio de la técnica de *webscraping*. Además no existe un *marco muestral formal*, conforme al universo, sin embargo se conocen la población total de links a investigar para detectar un dark pattern o no, en este sentido se considerará la fórmula de extracción de la muestra y considerar a la población infinita (cuando se desconoce el total de las unidades o la población supera a los 10.000 unidades muestrales)¹¹, la cual está representada por:

$$n = \frac{Z^2 * p * q}{e^2}$$

Donde;

n: Es el tamaño de muestra piloto, para detectar *Dark Patterns*

p: Probabilidad de éxito de encontrar la característica de estudio cualitativa, dentro de la población de estudio, para este caso, como se trata de una muestra piloto, se considerará la máxima probabilidad permisible de encontrar la característica de estudio que tomará el valor de 0,5

q: Probabilidad de no encontrar la característica de estudio de tipo cualitativa, dentro de la población en estudio, para este caso se trata de una muestra piloto se considerará la máxima probabilidad permisible para este parámetro de encontrar la característica de estudio el cual tomará el valor de 0,5 dado que es trata de $q = 1-p$


Z: Valor crítico de bajo la curva de la distribución normal estándar con media de 0 y desviación estándar de 1, en este caso, como se trata de muestra preliminar, se calculará con un nivel de confianza del 95% y nivel de significancia de $\alpha = 5\%$, considerando estos valores en las tablas de la distribución normal estándar el valor de *Z* tomará el valor de 1,96 que es el valor crítico dado los parámetros establecidos anteriormente.

e: Es el error de muestreo o nivel de precisión relativo, este está íntimamente relacionado con el intervalo de confianza deseado para en la determinación de este valor promedio de la variable en estudio, en este caso como se trata de una muestra piloto este valor tomará el valor de 5%, asumiendo normalidad.

Las posibles combinaciones de una muestra cualquiera está dado por:

¹¹ Fórmulas para el cálculo de muestras, SALUD EN TABASCO, Secretaria de Salud del Estado de Tabasco (pág 5) <https://www.redalyc.org/pdf/487/48711206.pdf>

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 27 de 25
	Detección de Dark Patterns por Web scraping		

$$S = C_{N,n} = (N \ n),$$

Entonces la probabilidad de que una unidad muestral sea extraída desde la población estará dada por:

$$P(s) = \frac{1}{(N \ n)}, \forall s \in S$$

Donde;

s : Es el tamaño de 1 sola muestra posible

S : Es el tamaño de todas las posibles muestras calculadas desde $C_{N,n}$

Con esta probabilidad se asegura que todas las posibles muestras de los links que se calculen y se seleccionan desde la población son equiprobables, es decir que cada muestra y cada unidad estadística tienen la misma probabilidad de ser seleccionadas.

Diseño muestral para (Empresas)

Para el estudio se realizó la selección de las empresas que se levantaron en *Dark Patterns* del año 2021, con lo que forma parte del juicio de la unidad investigadora UMM, y parte de esos criterios fue, seleccionar las empresas con mayores presencias detectadas de *Dark Patterns* en dicho estudio de, que son 37 empresas en total totalizando 11.916 casos de *Dark Patterns* en dicho estudio, el listado de empresas para el presente estudio es el siguiente:

RAZÓN SOCIAL	N° patrones oscuros
CENCOSUD RETAIL S.A.	567
CENTRO MÉDICO DE TRATAMIENTO AVANZADO DE LA PIEL LTDA	566
SERVICIOS DE VENTA INMOBILIARIA LIMITADA	557
DIRECTV CHILE TELEVISION LIMITADA	535
DIRECTV CHILE TELEVISION LIMITADA	534
RED CLINICA ESTETIKA MEDICA SPA	520
WIT INNOVACIÓN Y TECNOLOGÍA SPA	509
MERCADOLIBRE CHILE LTDA.	481
ANDACOR S.A.	472
IBERIA LINEAS AEREAS DE ESPANA S.A OPERADORA AGENCIA EN CHILE	439
EMPRESAS HITES S.A.	437
ASESORES EN VIAJES S.A.	425
COMERCIAL ECCSA S.A.	424
HIPERMERCADOS TOTTUS SA	410
SALCOBRAND S.A	403

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)


Detección de Dark Patterns por Web scraping

RAZÓN SOCIAL	N° patrones oscuros
DESPEGAR.COM CHILE S.A.	393
DESPEGAR.COM CHILE S.A.	392
SEGUROS FALABELLA	367
TRAVEL SECURITY S.A.	365
EASY S.A.	343
FARMAZON CHILE SPA	337
RAPPI CHILE SPA	335
EMPRESAS LA POLAR S.A.	319
DESARROLLO INMOBILIARIO MIPA S.A.	316
LATAM AIRLINES GROUP S.A.	262
CUPONATIC LATAM S.A	234
EXPORTADORA E IMPORTADORA PATRICIA CASTRO E.I.R.L.	233
ATRÁPALO CHILE S.A.	224
BIGFOOT CHILE SPA	194
BCI SEGUROS GENERALES S.A.	150
COMERCIALIZADORA SKECHERS CHILE LIMITADA	148
FRANCHISE CLINICA VITALIFE SPA	143
FALABELLA RETAIL S.A.	133
SODIMAC S.A.	132
MERCADOLIBRE CHILE LTDA.	127
CIA DE SEGUROS GENERALES CONSORCIO NACIONAL DE SEGUROS S A	126
CENCOSUD RETAIL S.A.	110
CENCOSUD RETAIL S.A.	109
LIDER DOMICILIO VENTAS Y DISTRIBUCION LTDA.	40
BUSCALIBRE S. A.	28
VIAJOBIEN SPA	25
SODIMAC S.A.	19

5.5) Aspectos metodológicos muestrales a considerar:

- **Universo:** Estará definido por la cantidad totales de links que se están en la webmaster de la empresa seleccionada.
- **Población:** Estará definida como todas las unidades estadísticas que son encontradas y extraídas desde los "sitemaps", con esto se delimita los sesgos no muestrales.
- **Unidad Estadística o unidad de observación:** Url's o links que forman parte de la población y tienen un cierto grado de probabilidad de ser seleccionados en la muestra.

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)

 SERNAC Servicio Nacional del Consumidor	Versión: 0.7	Estado: No Vigente	Página 29 de 25
	Detección de Dark Patterns por Web scraping		

- **Unidad Muestral:** Url's o links, que ya fueron seleccionados bajo un proceso aleatorio o no, en este caso es bajo un procedimiento o muestreo aleatorio.
- **Sesgo:** Error específico de la muestra por falta de representatividad, mayormente conocido como error total, este proviene de **error muestral** y el **error no muestral**.¹²
- **Error muestral:** Es el imputable al estudio de una parte de la población o muestra, para este caso en la muestra se representa como mencionamos anteriormente como "e"
- **Error no Muestral:** Es el que se produce en toda investigación como consecuencia definiciones conceptuales incorrectas, de fallos de instrumentos de medida, en la entrevista o en el desarrollo del área de trabajo.
- **Marco Muestral:** Normalmente es casi imposible confeccionar un marco muestral, que no excluya a algunas unidades estadísticas de la población, en nuestro caso, este listado, se confeccionará dado las url's resultantes o que se puedan directamente extraer desde los "sitemaps", de esta forma se delimitan los sesgos no muestrales en las unidades muestrales al momento de extraer la muestra. Para nuestro caso el marco muestral estará dado por el total de links que se puedan extraer del *sitemap* de cada página o dominio de las empresas seleccionadas.
- **Cobertura:** Empresas que fueron inspeccionadas en el estudio de *Dark Patterns* realizado por la UMM año 2021, mediante inspección *on - line*.

¹² Diseños Muestrales en la Investigación, Error, (pág 2 y 3), <https://webcache.googleusercontent.com/search?q=cache:5tymgXFwNf0J:https://dialnet.unirioja.es/descarga/articulo/5262273.pdf&cd=13&hl=es&ct=clnk&gl=cl>

	Elaboración	Revisión	Aprobación
Fecha	21.12.2022	22-12-2022	26-12-2022
Cargo	Profesional UMM (AMU/CVV/CMB)	Jefatura UMM (EMS)	Subdirector (GFZ)