

UNIVERSIDAD DE MÁLAGA

Programa de Doctorado en Tecnologías Informáticas



Modelado del proceso de compra del consumidor
final en los ecosistemas digitales mediante técnicas
de inteligencia computacional.

TESIS DOCTORAL

Francisco Elías Cabrera Lara

Director: Dr. José Ignacio Peláez Sánchez

Departamento de Lenguajes y Ciencias de la Computación





UNIVERSIDAD
DE MÁLAGA

AUTOR: Francisco Elías Cabrera Lara

 <https://orcid.org/0000-0001-6286-577X>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es



A mi familia



UNIVERSIDAD
DE MÁLAGA



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña FRANCISCO ELÍAS CABRERA LARA

Estudiante del programa de doctorado TECNOLOGÍAS INFORMÁTICAS de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: MODELADO DEL PROCESO DE COMPRA DEL CONSUMIDOR FINAL EN LOS ECOSISTEMAS DIGITALES MEDIANTE TÉCNICAS DE INTELIGENCIA COMPUTACIONAL.

Realizada bajo la tutorización de EZEQUIEL LÓPEZ RUBIO y dirección de JOSÉ IGNACIO PELÁEZ SÁNCHEZ (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 22 de NOVIEMBRE de 2021

<p>Fdo.: FRANCISCO ELÍAS CABRERA LARA Doctorando/a</p>	<p>Fdo.: EZEQUIEL LÓPEZ RUBIO Tutor/a</p>
<p>Fdo.: JOSÉ IGNACIO PELÁEZ SÁNCHEZ Director/es de tesis</p>	





UNIVERSIDAD
DE MÁLAGA

Dr. D. José Ignacio Peláez Sánchez, Catedrático de Universidad en el Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga.

Certifica:

Que D. Francisco Elías Cabrera Lara, Ingeniero en Informática por la Universidad de Málaga, ha realizado en el Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga el trabajo correspondiente a su Tesis Doctoral titulada:

Modelado del proceso de compra del consumidor final en los ecosistemas digitales mediante técnicas de inteligencia computacional.

Revisado el presente trabajo, estimo que puede ser presentado al tribunal que ha de juzgarlo, y autorizo la presentación de esta Tesis Doctoral en la Universidad de Málaga.

En Málaga, 19 de noviembre de 2021.

Fdo.: Dr. José Ignacio Peláez Sánchez
Director de la tesis
Catedrático de Universidad
Dpto. de Lenguajes y Ciencias de la Computación
Universidad de Málaga



UNIVERSIDAD
DE MÁLAGA

Agradecimientos

En primer lugar, a mi director de Tesis, José Ignacio Peláez Sánchez por haberme introducido al mundo de la investigación, por haber sabido guiarme, enseñarme y aconsejarme con sabiduría en todo este viaje y por haberme aceptado no solamente como miembro de su grupo de investigación sino como amigo.

En segundo lugar, a mi amigo y compañero Fabián, por haber sido la voz de la experiencia en toda esta aventura, estando siempre ahí presente para prestarme su ayuda siempre que la he necesitado.

A todos los miembros del grupo de investigación, por hacerme sentir el laboratorio como un segundo hogar para mí.

A mis amigos por su confianza, su ánimo, su comprensión y su paciencia a lo largo de estos últimos años.

Y por supuesto, a mis padres Paco y Lola y a mi hermana Nieves, por estar ahí siempre y darme todo el apoyo sin el cual este proyecto nunca hubiera sido posible.



UNIVERSIDAD
DE MÁLAGA

Índice de Contenidos

Agradecimientos	ix
Índice de Contenidos.....	xi
Índice de Figuras.....	xiii
Índice de Tablas	xv
Resumen	xvii
Abstract	xix
Capítulo 1: Introducción	1
1.1. La investigación de mercados	7
1.2. Efectos de la era de Internet sobre el comportamiento del consumidor.....	8
1.3. Comportamiento del consumidor.....	10
1.4. Las cinco fases de las transacciones comerciales.....	11
1.5. La importancia de los intangibles en el mundo empresarial	14
1.6. El modelo GE ² AN.....	15
1.7. La obtención de información acerca de los consumidores.....	18
1.7.1. La obtención de información solicitada	18
1.7.2. La obtención de información no solicitada	20
1.8. El análisis de la información de los ecosistemas digitales	25
1.9. Los indicadores de reputación online.....	27
1.10. La toma de decisión multicriterio.	29
1.10.1. Teorema de la imposibilidad de Arrow.....	30
1.10.2. Métodos de MCDM	32
1.10.3. El método AHP	33
1.10.4. El Eigenvector Method for Pairwise Voting (EMPV)	35
1.10.5. Teoría de conjuntos difusos	37
1.10.6. Operadores OWA.....	39
1.11. Cómo comprender la decisión del consumidor	41

1.11.1. La combinación de información solicitada y no solicitada	41
1.11.2. Las sinergias entre criterios	42
1.11.3. El valor en las comparaciones de los consumidores	45
1.12. Objetivos y estructura de la memoria.....	47
Capítulo 2: La combinación de información.....	49
Capítulo 3: Detección y cómputo automático de sinergias de criterios	65
Capítulo 4: Índice matricial geométrico de comparación: Un ORI construido sobre la comparación de alternativas.....	81
Capítulo 5: Conclusiones y líneas futuras	93
Chapter 5: Conclusions and future lines	97
Apéndice A: English summary	101
A.1. Consumer behavior and choice	103
A.1.1. The five stages of commercial transactions.....	103
A.1.2. Intangible assets management: the GE ² AN model.....	106
A.2. Information sources on consumer behavior	108
A.3. Online Reputation Indices	109
A.4. Multi Criteria Decision Making	110
A.4.1. Arrow's impossibility theorem.....	110
A.5. Factors affecting consumer decisions.....	112
A.5.1. The combination of solicited and unsolicited information	112
A.5.2. Synergies between consumer criteria	113
A.5.3. The value of comparisons.....	116
A.6. Conclusions	117
Bibliografía	121

Índice de Figuras

Figura 1: El proceso de decisión de compra.....	11
Figura 2: Entradas and salidas del proceso de toma de decisión realizado durante la fase de valoración de alternativas.....	13
Figura 3: Modelo de gestión de intangibles GE ² AN.....	16
Figura 4: Transformación del dato e información y conocimiento.....	22
Figura 5: Proceso de descubrimiento del conocimiento.....	23
Figura 6: Proceso ETL.....	24
Figura 7: Representación gráfica de los valores lingüísticos definidos en la tabla 2.....	39
Figura 8: Valoración de cuatro productos con respecto a “diseño”, “calidad” y “ahorro”....	43
Figure A1: The purchasing decision process.....	101
Figure A2: Inputs and outputs of the decision-making process performed on the alternatives valuation step.....	103
Figure A3: GE ² AN intangibles management model.....	105
Figure A4: Valuation of four products with regards to design, quality and price.....	112



UNIVERSIDAD
DE MÁLAGA

Índice de Tablas

Tabla 1: Valores de la escala de Saaty con sus significados.....	34
Tabla 2: Ejemplo de aplicación de TFN con 5 etiquetas lingüísticas.....	38



UNIVERSIDAD
DE MÁLAGA

Resumen

La accesibilidad de la información causada por el auge de los ecosistemas digitales ha provocado grandes cambios en las interacciones entre las empresas y consumidores. La oferta de productos y servicios se incrementa constantemente, lo que ha provocado que los consumidores sean cada vez más exigentes, estén más informados e interactúen más entre ellos. Esta interacción, por su parte, ha disparado el interés de las empresas en conocer cómo reaccionan los potenciales clientes a los productos y servicios ofertados, así como a la gestión por parte de la empresa de todos los aspectos relacionados con la cadena de producción.

Por ello, la apreciación de los públicos se ha convertido en una pieza clave para evaluar los activos intangibles de una empresa, disparando el interés de empresas e instituciones en que existan herramientas que permitan comprender el comportamiento del consumidor. Interés que viene motivado por la necesidad de conocer en qué se basan los públicos para tomar decisiones de compra y cuáles son las opciones que les resultan más atractivas de entre la enorme cantidad de productos que tienen a su alcance.

La clave para desentrañar las decisiones de compra de los consumidores reside, por tanto, en comprender el proceso de decisión subyacente a cada compra. Dentro de este proceso, una de las piezas clave es la evaluación de alternativas que se realiza en los momentos previos a la adquisición de un producto o servicio. En el campo de la inteligencia computacional, el modelado de esta decisión de compra utilizando técnicas de decisión multicriterio (MCDA) ya ha demostrado buenos resultados a la hora de comprender el proceso de toma de decisión del consumidor. Sin embargo, todavía hay una serie de problemas que quedan por resolver a la hora de comprender plenamente la decisión de los consumidores.

En este trabajo se han analizado los efectos que tienen estos problemas a la hora de modelar el comportamiento del consumidor a partir de información proveniente de los ecosistemas digitales y se han desarrollado nuevos procesos que resuelven cada uno de ellos.

En primer lugar, las diferencias entre ambos tipos de fuente de información empleadas dificultaban su integración, por lo que normalmente se trabajaba solo con parte de la información disponible. Esto introducía sesgos y hacía que fuera difícil obtener una imagen completa de lo que los públicos realmente sienten acerca de los productos y servicios. En este trabajo se presenta una metodología que permite integrar de manera sencilla ambos tipos de información y se ha demostrado además que el uso de información combinada mejora los resultados reduciendo la cantidad total de información necesaria.

En segundo lugar, los criterios que se empleaban para evaluar a los productos estaban siendo tratados como entidades independientes, cuando realmente se producían sinergias entre los mismos que no eran consideradas, reduciendo la precisión de los rankings de predicción de compra. El nuevo método de cálculo de sinergias, aportado en este trabajo, permite cuantificar de manera automatizable el valor de estas sinergias y emplearlos para el cálculo de un nuevo ranking, que considera dichas sinergias y por lo tanto, se ajusta mejor a los rankings reales de compra.

Por último, a la hora de cuantificar el valor apreciado de las alternativas, los indicadores de reputación online (ORIs) son la herramienta más utilizada, pero debido a la diferencia de escala interna entre distintos individuos y a que los ORIs tradicionales no están construidos en base al concepto de alternativas comparables, en el que los públicos se basan para evaluar las propias alternativas, el resultado de emplear estos ORIs como medida proxy del valor apreciado produce una serie de errores que reduce la validez de los rankings actuales de alternativas. Para reducir estos errores, en este trabajo se ha elaborado un nuevo ORI que se basa en agregaciones geométricas de matrices de comparación (CMGI). Este nuevo indicador se comporta de una manera análoga al proceso de evaluación que los consumidores llevan a cabo con las alternativas de compra y por lo tanto provee una medida más realista del valor percibido por parte de los consumidores de los productos en su contexto.

Aunque las tres metodologías planteadas en este trabajo pueden ser empleadas de manera independiente, su verdadero potencial aparece cuando se utilizan de manera combinada para modelar al completo la fase de evaluación de alternativas del consumidor, presente en todo proceso de compra. Todas las técnicas desarrolladas han sido probadas empleando datos reales provenientes de ecosistemas digitales y han demostrado su utilidad a la hora de mejorar la comprensión del proceso de compra del consumidor en los ecosistemas digitales.

Abstract

The accessibility of information caused by the rise of digital ecosystems has led to major changes in the interactions between companies and consumers. The supply of products and services is constantly increasing, which has led to consumers becoming more demanding, more informed and more interactive among themselves. This interaction, in turn, has triggered the interest of companies in knowing how potential customers react to the products and services offered, as well as to the company's management of all aspects related to the production chain.

For this reason, the appreciation of the public has become a key element in evaluating the intangible assets of a company, triggering the interest of companies and institutions in the existence of tools that make it possible to understand consumer behavior. This interest is motivated by the need to know what audiences base their purchasing decisions on and which alternatives are the most attractive among the myriad of products available.

Therefore, the key to unraveling consumers' purchasing decisions lies in understanding the decision process underlying each purchase. Within this process, one of the key pieces is the evaluation of alternatives that takes place in the moments prior to the purchase of a product or service. In the field of computational intelligence, the modeling of this purchase decision process using multi-criteria decision making techniques (MCDA) has already shown good results in understanding the consumer's decision making process, both from solicited and unsolicited information. However, there are still a number of problems that remain to be solved in order to fully understand the consumer decision.

In this work, we have analyzed the effects of these problems in modeling consumer behavior based on information from digital ecosystems and developed new processes that solve each of them.

Firstly, the differences between the two types of information sources used made it difficult to integrate them, causing only part of the available information to be used. This introduced biases making it difficult to obtain a complete picture of what audiences really feel about products and services. This work presents a novel methodology that makes it possible to integrate both types of information in a simple way and has also shown that the use of combined information improves the results by reducing the total amount of information needed.

Secondly, the criteria used to evaluate the products were being treated as independent entities, when in fact there are synergies between criteria that were not considered, reducing the accuracy of purchase prediction rankings. The new method for calculating synergies provided in this work makes it possible to automatically quantify the value of these synergies and use them to calculate a new ranking, which takes these synergies into account and is therefore better adjusted to the real purchase rankings.

Finally, when it comes to quantifying the perceived value of alternatives, online reputation indicators (ORIs) are the most widely used tool, but due to the difference in internal scales between different individuals and the fact that traditional ORIs are not built around the concept of comparable alternatives, on which audiences themselves rely to evaluate the alternatives, the result of using these ORIs as a proxy measure of perceived value produces a series of errors that reduce the validity of the current rankings of alternatives. As part of this work, a new ORI has been developed that is based on geometric aggregations of comparison matrices (CMGI). This new indicator behaves in a manner analogous to the evaluation process that consumers undertake when purchasing alternatives and therefore provides a more realistic measure of how consumers perceive the value of products within their context.

Although the three methodologies presented in this paper can be employed independently, their true potential appears when used in combination to fully model the consumer's alternatives evaluation phase, which is present in every purchase process. All the techniques developed have been tested using real data from digital ecosystems and have proven their usefulness in improving the understanding of the consumer buying process in digital ecosystems.

Capítulo 1: Introducción

Los ecosistemas digitales han cambiado la forma en la que se comunican las personas. Antiguamente, la comunicación masiva y pública estaba limitada a medios de comunicación de manera unidireccional a través de los medios impresos, la radio y la televisión, mientras que los particulares eran meros consumidores de contenido y la comunicación entre los mismos era principalmente efímera y cara a cara. Con la aparición de los medios sociales los consumidores adquirieron la posibilidad de realizar comunicaciones de carácter público, convirtiéndose prácticamente de la noche a la mañana en generadores de contenidos, además de seguir manteniendo su rol como consumidores de estos.

Este cambio de paradigma ha dejado atrás los modelos lineales de la comunicación, como la teoría de la Bala Mágica [1]. Ya que no solo aumentaron las interacciones entre los propios públicos, sino que estos pueden generar *feedback* y opiniones e incluso en última instancia convertirse en *prosumidores*, es decir, individuos capaces de ser productores y consumidores de contenido, un concepto que introduce la interactividad de los públicos.

Este cambio en la forma de interactuar con los públicos ha supuesto un gran salto evolutivo, no solo para la sociedad, sino para el modo en el que empresas e instituciones se comunican, ya que el alcance de las opiniones expresadas es cada vez mayor y una opinión desfavorable con una alta repercusión tiene el potencial de causar una reacción en cadena con graves consecuencias para la reputación de alguno de sus activos, por lo que los proveedores de productos y servicios no pueden permitirse el lujo de ignorar las opiniones que se vierten continuamente acerca de ellos en los ecosistemas digitales, prestando especial atención a indicios que puedan ayudar a detectar crisis de forma temprana.

Cuando se produce cualquier tipo de crisis en una organización, esto supone una ruptura en el devenir institucional, generando así consecuencias para todos los elementos que trabajan y rodean a la organización. Esta afectará en mayor o menor medida tanto a los públicos internos como a los externos, lo que implicará consecuencias para su identidad y su imagen.

La gestión de una crisis en materia de comunicación es un trabajo arduo que conlleva un seguimiento de fondo, pero que genera grandes beneficios en forma de intangibles. Contar con una identidad en la que los públicos internos, entre los que encontramos trabajadores y accionistas, se sientan cohesionados con la organización, es imprescindible para el devenir de la empresa, al igual que lo es la gestión de la percepción externa en forma de la imagen percibida por parte de clientes potenciales, medios de comunicación y demás públicos externos.

La problemática reside en que, si bien trabajar estos aspectos lleva años, una crisis puede producir que la falta de confianza se instaure en los públicos y perder en minutos el trabajo de fondo llevado a cabo durante décadas. Por ello, es primordial que las organizaciones cuenten con estrategias, protocolos y sistemas de detección preparados de antemano para la gestión de crisis. De esta manera, se puede, en el momento que estalle una crisis y estemos en el foco de los medios de comunicación, reducir el impacto negativo hacia nuestra reputación.

Tratar una crisis es algo que hay que trabajar desde dentro de la organización e identificarla lo antes posible es el primer paso para una buena gestión. Cualquier comentario o crítica que ataque a la esencia de la organización podría convertirse en una crisis potencial y más aún en un momento en el que la viralidad en los medios sociales está a la orden del día. Por ello aprovechar las herramientas de las que disponemos y crear otras nuevas que nos permitan monitorizar las opiniones de los públicos puede ser clave para una actuación precisa, eficaz y temprana, evitando de esta forma daños estructurales en la imagen e identidad de la organización.

Además, las empresas difícilmente existen como un ente aislado en un monopolio, al contrario, en los últimos tiempos se ha incrementado notablemente la competencia que experimentan muchas marcas debido a la accesibilidad derivada del proceso de globalización que estamos viviendo. Hace 50 años era relativamente difícil acceder a productos provenientes de otros países y, sin embargo, en la actualidad millones de usuarios de Internet compran productos que son enviados desde otro continente, en muchos casos con periodos de entrega inferiores a una semana. Con lo que al igual que la comunicación ha dejado de estar restringida a aquellos con quienes teníamos contacto directo, las opciones de consumo también han relajado esta limitación.

Teniendo en cuenta este entorno, nos encontramos dos casuísticas que pueden retroalimentarse entre sí. Por un lado, los usuarios cada vez más activos e interrelacionados, hacen públicas sus opiniones, inquietudes y deseos, siendo esto una fuente inagotable de información.

Por otro lado, la competencia en nuestros días se ha vuelto más feroz que nunca y como consecuencia, cualquier ayuda que podamos emplear como ventaja competitiva será bienvenida por las empresas. El atender a la información que los públicos vierten en ecosistemas digitales es sin duda una de estas ventajas, ya que provee opiniones de primera mano de la aceptación que tienen los productos y servicios que las empresas ofrecen, información acerca de la apreciación que se tiene de la misma con respecto a su competencia, así como datos acerca de los puntos que parecen más o menos atractivos de las marcas que se les ofrecen a los públicos.

La comunicación bidireccional, por lo tanto, ha generado una suerte de retroalimentación continua entre negocios y público, en la que las opiniones de los públicos generan un efecto en las decisiones que toman las empresas y las consecuencias de estas decisiones a su vez son recibidas por los públicos, provocando la generación de nuevas opiniones al respecto que alimentan este bucle de comunicación.

Para la mayoría de los comunicadores, este gran cambio en la forma de interactuar con las personas ha supuesto una evolución tanto en la forma de tratar la información como en el modo de gestionar las comunicaciones de la empresa. Hace unos pocos años los responsables de comunicación eran los que, a través de los medios tradicionales, brindaban la información. Esto ha quedado muy lejos del paradigma actual, en el que los responsables de comunicación han pasado de ser una de las pocas fuentes de comunicación disponibles a convertirse en una más de las voces que pueden oírse constantemente en un mar de información, con la responsabilidad añadida de tratar de controlar el discurso de forma que se mantenga una reputación y una apreciación de la marca lo más positiva posible. Es por ello, por lo que la comunicación y el análisis de estas fuentes ha adquirido en los últimos tiempos una mayor importancia a la hora de tomar decisiones dentro de las empresas e instituciones.

En el ámbito empresarial, las decisiones estratégicas habitualmente se han tomado a través del análisis de los datos históricos disponibles [2], pudiendo entre otras cosas analizar las causas de un problema o situación actual, determinar las estrategias posibles y predecir el resultado de aplicar dichas estrategias. Durante mucho tiempo, los datos históricos utilizados para el análisis eran principalmente datos internos provenientes de sistemas de información transaccionales como registros de clientes, ventas, compras, empleados, etc. Sin embargo, el auge de la sociedad de la información ha permitido diseñar nuevos tipos de análisis a partir de los datos generados en los ecosistemas digitales.

Los análisis de estos datos se han convertido en herramientas esenciales para el buen funcionamiento de las empresas, las cuales están integrándolos cada vez más y de manera transversal en todos los eslabones de su cadena productiva. Como consecuencia, la información acerca de la apreciación de los públicos sobre la marca y el sector es ahora una de las métricas más importantes para evaluar el desempeño actual de corporaciones e instituciones, además de ser una pieza clave para decidir las futuras acciones que las empresas deberán llevar a cabo en el futuro [3].

Por lo tanto, nos enfrentamos ante el reto de transformar cantidades masivas de datos en conocimientos valiosos que puedan emplearse para desarrollar estrategias enfocadas en la toma de decisiones. Esta es una tarea que requiere la localización, extracción y tratamiento de estos datos de manera que estos se puedan presentar de la forma más adecuada en cada

momento, facilitando así su uso y empleabilidad, a pesar de que estos datos puedan provenir de fuentes muy diferentes entre sí. Y es que entender e interpretar la información digital exige comprender la complejidad del contexto en el que surge y se desarrolla dicha información.

Es por ello que en los últimos tiempos ha aumentado la demanda de herramientas que permitan entender las capacidades situacionales de las organizaciones y sus marcas. De manera que se puedan ofrecer productos realmente atractivos, que sean capaces de generar experiencias, emociones y actitudes de marca consistentes y que puedan ser una opción convertible en negocio.

Aunque existen múltiples modelos y herramientas para evaluar el estado actual de las opiniones de los públicos, los negocios demandan el ir un paso más allá y contar con herramientas que sean verdaderamente capaces de integrarse en los procesos de toma de decisiones de la empresa, de una manera que la información obtenida pueda ser realmente útil.

Es decir, no solo centrarse en analizar esa información, sino emplearla para obtener respuestas más efectivas para la toma de decisiones. Y es que, las grandes empresas ya se encuentran en un punto en el que son capaces de medir correctamente la dirección y la fuerza del viento que sopla en los medios digitales, pero lo que verdaderamente necesitan saber es la orientación de las velas necesaria para llegar lo más lejos posible.

Para ello, es necesario el desarrollo de nuevas técnicas que permitan simular conceptos, perfiles y segmentos que ayuden a comprender los comportamientos de la sociedad digital en diferentes contextos, mientras se realiza una vigilancia de las opiniones vertidas en los ecosistemas digitales. Esta vigilancia nos permite estar atentos a escenarios que puedan suponer un riesgo para el valor de la marca y así poder adelantarnos a situaciones que perjudiquen a la empresa.

El éxito de herramientas comerciales como Brandwatch [4], Keyhole [5], [6] o Conviva [7] que permiten monitorizar y medir campañas sociales es una prueba de la gran demanda de este tipo de metodologías por parte de empresas, instituciones y organismos públicos, que necesitan actualizar de manera continua su conocimiento competitivo sobre las dinámicas y tendencias que van imponiéndose en cada momento debido a las frecuentes interacciones con su público [8].

Sin embargo, aunque el disponer de información proveniente de ecosistemas digitales es un paso importante para las decisiones empresariales, el objetivo de fondo es conocer con el mayor detalle posible la situación del mercado dentro de una industria, por lo que esta información no es más que una pieza más dentro del contexto de la investigación de mercados.

Si realmente se desea obtener conocimiento que pueda resultar útil para comprender al consumidor en su contexto, es necesario modelar su comportamiento para extraer los patrones que influyen en la decisión de compra implícitos en los datos. Para ello, son varios los problemas que se plantean a resolver en este trabajo encaminados a este fin.

Por un lado, es necesario combinar los distintos tipos de información disponible en los ecosistemas digitales en forma de información solicitada y no solicitada para obtener una imagen global, más completa, de lo que los consumidores opinan acerca de marcas y productos. Por otra parte, aunque ya existen técnicas que permiten extraer parte de la información necesaria, esta se trata de manera independiente sin tener en cuenta las sinergias que se producen entre las características de los productos, por lo que se hace necesario desarrollar una metodología que permita identificar y procesar estas sinergias. Y, por último, los indicadores reputacionales online que se emplean en la actualidad como medidas proxy del valor de los productos presentan serias limitaciones derivadas de su concepción original como elementos para valorar aisladamente dichos productos que deben ser atendidas.

Por lo tanto, en este trabajo se plantean los siguientes tres objetivos encaminados a la obtención de conocimiento acerca del proceso de compra del consumidor final en los ecosistemas digitales:

1. Resolver el problema de la combinación de información solicitada y no solicitada.
2. Resolver el problema de la detección e integración de las sinergias entre criterios de compra.
3. Mejorar la fiabilidad de los indicadores de reputación online a la hora de que estos puedan ser utilizados como medida proxy del valor percibido por los consumidores de los productos.

Para presentar solución a los problemas anteriormente descritos, se ha estructurado este trabajo de la siguiente forma: En el capítulo de introducción se presenta la situación actual en cuanto al modelado del consumidor en los ecosistemas digitales, su contexto dentro de la investigación de mercados y se explican en detalle los efectos que causan los problemas anteriormente mencionados. En el segundo capítulo se plantea una solución al problema de la combinación de información solicitada y no solicitada gracias a la inclusión de información de rankings obtenidos mediante métodos de votación por pares en cuestionarios tradicionales realizados con escalas de Likert. En el tercer capítulo se presenta una solución determinista y automatizable al problema de la detección y procesamiento de sinergias de criterios empleando una combinación de técnicas de toma de decisión multicriterio pertenecientes a la familia AHP, herramientas provistas por la teoría de lógica difusa y el uso de la integral de Choquet como método de agregación. En el cuarto capítulo se introduce un nuevo índice que mejora la medición en el valor percibido de los productos directamente a partir de información proveniente de los ecosistemas digitales basado en agregaciones geométricas de matrices de

comparaciones, que es capaz de extraer de manera directa información de preferencias a partir de comparaciones expresadas en lenguaje natural. Finalmente, en el capítulo 5 se exponen las conclusiones que derivan de esta memoria, un resumen de los principales aportes y las líneas de trabajo futuras abiertas.

1.1. La investigación de mercados

La investigación de mercados [9] consiste en la recolección y análisis de los datos acerca de la situación del mercado dentro de su industria. Las compañías necesitan estos datos para crear nuevos productos y servicios que puedan ser atractivos a los consumidores, dirigiendo la toma de decisiones con conocimiento acerca de los públicos objetivos. Se trata de un aspecto importante ya que la gestión de la demanda de productos y servicios es una de las principales bases de un negocio próspero; además de para implementar mejoras en productos ya existentes y adaptarlos a las cambiantes demandas del mercado.

Ya sea para la creación de un nuevo negocio, el lanzamiento de un nuevo producto o el desempeño del día a día, siempre ha sido importante para las empresas comprender el comportamiento de sus clientes. Los consumidores demandan soluciones para sus necesidades, de manera que el principal objetivo de la investigación de mercados es comprender el comportamiento de estos para así ofrecerle bienes y servicios que puedan ser percibidos como útiles para la satisfacción de sus necesidades.

Desde el punto de vista del consumidor, existen muchos tipos de problemas cuya solución puede satisfacerse mediante la adquisición de algún bien o servicio, desde comprar una herramienta que le permita facilitar el desempeño de alguna tarea, hasta obtener entretenimiento para sus momentos de ocio, pasando por realizar una inversión o incluso aparentar un mayor estatus social simplemente por la posesión de bienes de determinadas características o marcas. Algunas de estas motivaciones son más obvias que otras, aunque al final es necesario que los productos disponibles permitan atajar una necesidad, por muy concreta que esta sea, ya que si un producto es percibido como inútil en todos sus aspectos no existirá ningún tipo de motivación que pueda llevarlos a la adquisición del producto, por lo que este difícilmente será adquirido por los consumidores.

La información es, por lo tanto, clave para un negocio exitoso ya que permite no solo conocer las necesidades de los públicos actuales sino también de los potenciales; además sirve como una ventana para observar el estado de nuestra propia competencia, así como la oferta actualmente disponible y la respuesta que esta ha tenido a lo largo del tiempo por parte de los consumidores. Estos datos nos permiten evaluar las decisiones de gestión con el fin último de adaptar el comportamiento corporativo para proveer experiencias, de acuerdo con lo que los consumidores perciben como las características más atractivas en los productos o servicios disponibles.

Las técnicas empleadas en la investigación de mercados tienen múltiples aplicaciones en la gestión corporativa, tales como el desarrollo de productos, la publicidad, la investigación de marcas, el *concept-testing*, el cálculo de precios y la exploración de la satisfacción de los consumidores. Cada uno de estos aspectos aporta valor a las compañías que invierten recursos en ellos.

Un claro ejemplo lo vemos en cómo la publicidad es capaz de crear o mejorar la percepción de necesidad que distintos grupos de consumidores tiene asociada a un determinado producto, lo que en las últimas décadas ha convertido a las campañas publicitarias en un pilar fundamental de las actividades de marketing de las compañías, llegando en ocasiones a superar el coste de desarrollo del propio producto ofertado. Esta inversión se realiza porque la publicidad ha demostrado su impacto positivo en el valor intangible de las empresas y su capacidad para incrementar el valor de las marcas [10], [11]. Sin embargo, para conseguir el máximo rendimiento posible de técnicas de marketing tales como la publicidad, se hace necesario un gran conocimiento acerca de los clientes potenciales y objetivos. De manera más concreta, necesitamos comprender el proceso de decisión del consumidor, es decir, cómo y por qué los consumidores eligen unos productos y servicios específicos para tener una base que nos permita establecer nuestros objetivos.

Las fases de este proceso están ya definidas en la literatura científica desde el punto de vista del negocio y cuentan con un marco teórico que puede servir de base para llegar a comprender cómo actúan los consumidores, pero no existe todavía ningún modelo completo que sea capaz de explicar el comportamiento de los consumidores en ecosistemas digitales [12] de forma que se pueda integrar esta información en un proceso de toma de decisiones. Sin embargo, como paso previo a este modelado, es necesario comprender en detalle la influencia actual que producen los medios digitales como fuente de información en el consumidor, antes de poder considerarlas como un elemento clave de dicho modelo.

1.2. Efectos de la era de Internet sobre el comportamiento del consumidor

Indudablemente, Internet ha cambiado las interacciones entre las empresas y los consumidores, las economías de prácticamente todos los países han evolucionado muy rápidamente hacia los ecosistemas digitales. Este cambio ha afectado no solamente a la economía, sino a otras áreas

de la sociedad tales como la educación, la ciencia y la innovación, haciendo del mundo un lugar mucho más interconectado.

Bajo estas nuevas condiciones, las asociaciones con científicos, universidades y proveedores de servicios externos se han tornado una necesidad para muchos sectores industriales y la investigación de mercado ha pasado de ser el trabajo previo a realizar por algunas compañías que se lo podían permitir a ser un aspecto importante en cada una de las fases del desarrollo y comercialización de productos. La accesibilidad a la información estimula la actividad innovadora y reduce costes, incrementando las posibilidades de obtener beneficios en un mercado floreciente y bastante atractivo.

Además, este efecto funciona a la inversa, ya que, en una sociedad altamente conectada con los consumidores, los mismos desean saber tanto sobre las empresas como las empresas acerca de los consumidores. Hay una preocupación real por parte de los públicos sobre la forma en la que las empresas desempeñan sus tareas, cómo tratan a sus empleados, cómo gestionan los efectos medioambientales y sociales derivados de su producción y otros aspectos similares. En este sentido, aspectos como la transparencia y la coherencia de las comunicaciones se han convertido en temas muy importante para los clientes y en algunos sectores ya se ha demostrado que estos están estrechamente ligados con los beneficios. [13]–[15].

Sin embargo, el poder que hoy en día tienen los ecosistemas digitales va mucho más allá de la influencia en la apreciación de marca o las ventas de los productos patrocinados. Los ecosistemas digitales afectan a todos los ámbitos de la empresa, desde la capacidad para atraer y retener empleados hasta el valor en bolsa de las compañías. Un claro ejemplo de esto último fue la crisis de GameStop de principios de 2021 [16], en la que un grupo de usuarios de la red social Reddit compartió información acerca de cómo algunos fondos financieros podrían estar haciendo múltiples operaciones bursátiles en posición corta con acciones de la compañía GameStop (NYSE: GME). La recomendación de compra de estos usuarios fue escuchada por una enorme cantidad de usuarios de la red social que, actuando como pequeños inversores, comenzaron a comprar acciones de la compañía de manera masiva, llegando a provocar una subida superior al 1.500% en el valor de las acciones en menos de 15 días [17]. Casos como este ponen de manifiesto el poder que a día de hoy tienen tiene la información disponible de manera abierta en Internet y el enorme grado de digitalización que tiene la sociedad actual.

Una de las consecuencias más directas de este proceso de digitalización es que, dado que la información se ha convertido en un importante activo económico, la sociedad está virando hacia una economía cada vez más basada en el conocimiento, donde uno de los principales factores para el éxito de toda la cadena de producción es precisamente un buen entendimiento de todas las piezas involucradas en la misma [18]. Por este motivo, las empresas están cada

vez más asignando una notable cantidad de sus recursos a la comprensión del comportamiento de los consumidores en la era digital, para así tratar de desentrañar el proceso mental de los clientes potenciales a la hora de elegir los productos y servicios que adquieren y así aplicar este conocimiento a la gestión de los bienes tanto tangibles como intangibles de la empresa [19], [20].

1.3. Comportamiento del consumidor

Uno de los objetivos más claros de las empresas que invierten en estudios de mercado es precisamente comprender el comportamiento que los potenciales clientes exhiben al buscar, evaluar, decidir y comprar productos servicios disponibles en el mercado que puedan satisfacer sus necesidades. A la hora de comprender este comportamiento, un aspecto fundamental es reconocer que los recursos disponibles para el consumidor, como por ejemplo el dinero y el tiempo disponibles para cada individuo, son limitados. Esta disponibilidad de recursos restringe la capacidad de los clientes potenciales para obtener soluciones ofertadas, lo que provoca un problema de decisión cada vez que un consumidor tiene una necesidad que pueda ser satisfecha por algún producto o servicio que debe ser estudiado.

Como cualquier problema de decisión, el comportamiento del consumidor se puede modelar como un proceso cognitivo [21]. En este proceso, el mismo consumidor recibe una serie de inputs en forma de estímulos provenientes de su entorno que, compitiendo por su atención, se unirán a la información subjetiva presente en la memoria del individuo proveniente de experiencias y emociones previas, que posteriormente son procesadas mediante un proceso de toma de decisión multicriterio que el consumidor realiza de manera natural. Las creencias, metas y otros aspectos intrínsecos de los consumidores que varían de un individuo a otro afectan a este proceso, causando diferentes comportamientos incluso aunque estos sean expuestos a los mismos estímulos. El resultado de este proceso es la actitud del consumidor hacia el producto que finalmente se traduce en la decisión final de compra [22].

1.4. Las cinco fases de las transacciones comerciales

La clave para desentrañar el comportamiento del consumidor reside en comprender el proceso de decisión subyacente a cada compra. Este proceso consta de las 5 fases que se muestran en la figura 1, que son: Reconocimiento de una necesidad, búsqueda de información, evaluación de las opciones disponibles, decisión de compra y comportamiento post-compra [23]. Estos cinco pasos son la base del razonamiento que cualquier persona realiza de manera subconsciente al momento de adquirir un bien o servicio y su comprensión es esencial tanto para entender las intenciones del consumidor antes y después de la compra, como para crear satisfacción y lealtad hacia la marca.

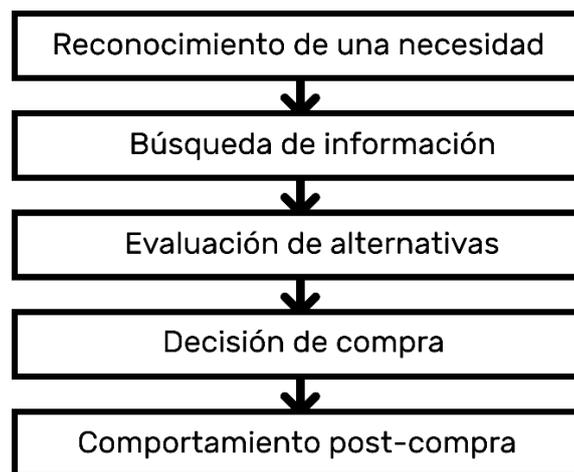


Figura 1: El proceso de decisión de compra

El proceso de compra comienza con el reconocimiento de una necesidad por parte del potencial consumidor. Este paso se realiza generalmente de manera inconsciente y puede incluir diversas motivaciones dependiendo del tipo de producto a ser comprado. Estas motivaciones pueden abarcar desde razones prácticas y sencillas de comprender, como el hambre o las necesidades de transporte, a razones más subjetivas y complejas, como la necesidad de ser percibido de cierta manera por otras personas o el experimentar una sensación de logro personal. Dado que es precisamente la percepción de una necesidad la que inicia todo el proceso de compra, existe un gran interés en la literatura científica acerca de cómo identificar o crear dichas necesidades,

así como técnicas empleadas a lo largo del tiempo para alterar de manera exitosa las percepciones que hay respecto a estas necesidades [24]–[26].

Una vez el consumidor ha identificado una necesidad a ser satisfecha, el segundo paso del proceso de decisión de compra es la búsqueda de información. Esta búsqueda puede ser de dos tipos. Por un lado, tenemos la percepción pasiva, en la que los consumidores se limitan simplemente a ser perceptivos a los estímulos recibidos de su entorno a través de diferentes fuentes, y por otro la búsqueda activa, en la que los consumidores realizan acciones encaminadas a la búsqueda de una solución a su necesidad reuniendo información acerca de productos y servicios que potencialmente puedan cubrir sus necesidades. Esta información incluye tanto datos objetivos, tales como características técnicas de un producto como información subjetiva, como pueden ser las opiniones de otros usuarios de productos que cubran necesidades similares.

Una fuente de uso mayoritario a la hora de obtener opiniones de productos son las conversaciones “boca a boca” entre consumidores, estas están caracterizadas por ser una forma de comunicación directa entre receptor y comunicador y no son percibidas como comunicaciones comerciales hacia una marca, servicio o producto por parte del receptor. Este tipo de comunicaciones han sido muy estudiadas desde la década de los 50, pero con el auge de las tecnologías digitales y especialmente de los social media, el concepto de lo que se consideraba “boca a boca” se ha ampliado notablemente. Décadas atrás, los consumidores tenían acceso como mucho a un par de docenas de opiniones acerca de productos o marcas que fueran populares; y en el caso de muchos productos de nicho o menos populares era habitual no conocer a nadie con experiencias que pudieran compartir, sin embargo, con el acceso a Internet, todo el mundo puede interactuar de manera instantánea con miles de opiniones acerca de prácticamente cualquier producto [27].

Esta información, cuando se utiliza correctamente, provee de muchísimas oportunidades que las empresas pueden aprovechar. En primer lugar, los consumidores dejan de estar restringidos a aquello que está disponible en su entorno geográfico, debido a la capacidad de realizar compras online. Desde el punto de vista de la empresa, esto implica que sus clientes potenciales ya no se encuentran en el mismo barrio o la misma ciudad donde se encuentra su comercio, sino que, cuando se aprovechan correctamente las nuevas tecnologías, cualquier persona del planeta puede convertirse en un cliente. En segundo lugar, dado que la información de opinión vertida en Internet suele tener un carácter público, esta provee a los negocios información acerca de cómo los consumidores perciben el producto [28].

El tercer paso del proceso de decisión del consumidor es la valoración de alternativas. En esta fase, el consumidor hace uso de la información que ha recopilado en la fase anterior para

determinar cuál de los productos, si es que hay alguno, es el más apropiado para atajar sus necesidades. Con este propósito, el consumidor prospectivo evalúa las diferentes características de las alternativas a su alcance de acuerdo con la información que este tiene en su memoria acerca de experiencias, creencias, actitudes, metas, poder adquisitivo y cualquier otro aspecto que sea capaz de recordar. Tal y como se muestra en la figura 2, este paso se puede modelar como un proceso de decisión multicriterio (del inglés *Multi-Criteria Decision Making* - MCDM) si consideramos las características del producto como las alternativas en las que los consumidores se apoyan para tomar una decisión. El ranking resultante de este proceso pasa entonces a ser el punto de partida del siguiente paso del proceso de decisión de compra.

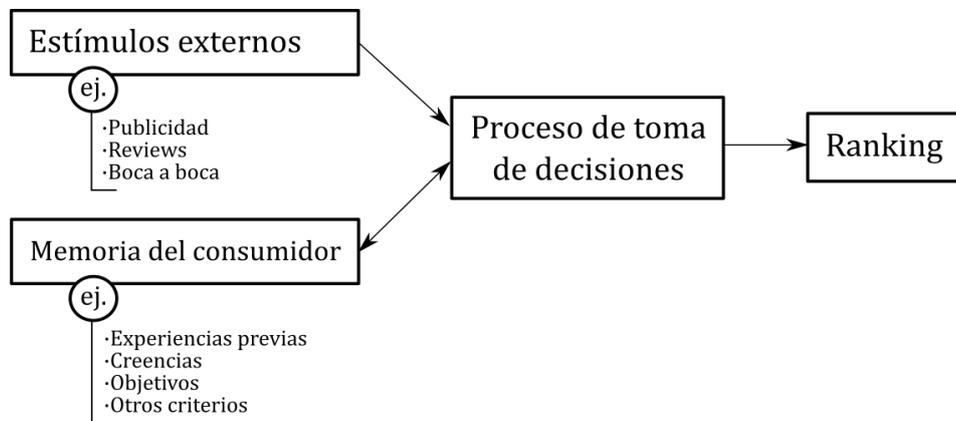


Figura 2: Entradas and salidas del proceso de toma de decisión realizado durante la fase de valoración de alternativas

En la cuarta fase del proceso de decisión de compra, el consumidor realiza la acción más apropiada de acuerdo con el ranking que ha establecido. Existen múltiples factores externos que pueden afectar a su acción a la hora de llevar a cabo la misma como pueden ser la disponibilidad, cambios en el precio del producto a la hora de realizar la compra o los plazos de entrega. No solo eso, sino que además las acciones que el consumidor puede realizar no son binarias, en el sentido de que este puede, además de proceder o no a la compra, optar por un préstamo o alquiler, decantarse por una solución sustitutiva para su problema o aplazar la decisión de compra entre otras opciones.

El último paso del proceso de decisión de compra es el comportamiento post-compra que se lleva a cabo una vez realizada, o no, la compra. Este paso es de suma importancia porque afecta no solo a los futuros hábitos de compra del consumidor que ha participado de la

transacción comercial, sino que las experiencias de este consumidor pueden influir también en otras personas que se vean expuestas a comunicaciones emitidas por el mismo. En esta fase, la propia acción de acometer un proceso de compra genera nuevas emociones, que son retenidas en la memoria de los consumidores y se transformarán en una nueva experiencia que podrá ser sopesada en futuras compras.

No obstante, dado que estas emociones pueden ser fácilmente transmitidas tanto de manera tradicional como aprovechando las nuevas tecnologías, cualquier otro consumidor prospectivo puede adquirir las experiencias generadas en esta fase por multitud de consumidores e incorporarlas a su memoria en la fase de búsqueda de información de un proceso de decisión de compra posterior. Debido a esto, cualquier compra puede potencialmente influir en todas las compras que se lleven a cabo en un momento posterior, por lo que la gestión de lo que los clientes piensan acerca de los productos adquiridos es una parte importante de la gestión de intangibles que tienen las empresas en la actualidad.

1.5. La importancia de los intangibles en el mundo empresarial

Dentro de los bienes que habitualmente gestionan las empresas nos encontramos dos categorías claramente diferenciadas. Los bienes tangibles son aquellos elementos, generalmente físicos, con un valor monetario determinado que puede ser fácilmente inventariar y cuantificable. Los edificios, maquinaria, vehículos y material informático que formen parte del patrimonio de la empresa son un claro ejemplo de bienes tangibles. Por otro lado, los bienes intangibles son aquellos que, aunque también pueden ser cuantificables a nivel monetario, son más difíciles de identificar y medir porque no vienen con el precio marcado. La lealtad hacia la marca, la reputación del producto o la apreciación del riesgo que tiene la propia compañía en el mercado financiero son elementos que, aunque no son físicamente tangibles, tienen una enorme influencia en la capacidad de las empresas para obtener beneficios.

De la misma manera que cada vez es más sencillo y rápido el acceso a la información por parte de los *stakeholders*, la importancia que tienen los activos intangibles de las empresas también ha ido creciendo. Experiencias relacionadas con variables intrínsecamente intangibles, como son la responsabilidad social corporativa y el comportamiento ético apreciado de una empresa, tienen un efecto claro en el valor de esta en el mercado de valores. Prueba de ello es que, en

momentos de crisis, una disculpa por parte de una compañía directamente responsable de la misma mitiga las pérdidas debidas a la crisis, y sin embargo no disculparse o hacerlo sin haber sido el causante afecta negativamente al valor de la empresa [29].

El efecto que producen los activos intangibles en el valor económico de una empresa depende además de una multitud de factores, esto provoca que lo que para una empresa puede funcionar perfectamente, para otra puede causar pérdidas millonarias en el valor de sus acciones en bolsa. Entre múltiples factores, el hecho de que una empresa sea conocida o apreciada por el público afecta enormemente a cómo debe gestionarse la reputación corporativa de la misma. Una empresa muy conocida notará más los efectos de una mala comunicación empresarial en un momento dado, sin embargo, en estos casos el efecto es muchísimo menor si la empresa goza de una buena reputación entre sus *stakeholders* [30].

Todo esto ha hecho que los intangibles hayan pasado a un primer plano en lo que a gestión empresarial se refiere. Las compañías, se han dado cuenta del potencial que un manejo de estos activos tiene para proveer un valor diferencial en un mundo cada vez más competitivo, por lo que estas en los últimos años han venido asignando una mayor cantidad de recursos a su estudio y análisis. Hoy día ya no se trata simplemente de un lujo que se pueden permitir algunas corporaciones, ya que un buen manejo de los intangibles es capaz de reducir costes, atraer inversores, incrementar el valor bursátil, afianzar lealtad de los consumidores, atraer y retener mejores empleados y minimizar el impacto de algunas crisis [31], [32]. Por ello, las compañías tienen una necesidad real de conocer la percepción que los públicos tienen acerca de ellos antes de poder ajustar sus estrategias de negocio [33].

1.6. El modelo GE²AN

La forma de conseguir esta gestión del capital intangible es la construcción de sistemas capaces de proveer un mejor entendimiento de estos valores para apoyar los procesos de decisión que habitualmente se llevan a cabo en las posiciones de gestión de las empresas. Para que estos sistemas puedan cumplir bien su función, deben ser capaces de extraer conocimiento de grandes fuentes heterogéneas de datos, tanto internas como externas a la empresa, y procesar esta información de manera conjunta para proveer un correcto entendimiento del modelo de negocio y su relación con los consumidores. Los modelos empleados para esta tarea deben ser lo suficientemente generales como para abarcar información proveniente de cualquier fuente de datos según su grado de pertinencia con respecto a la empresa para la que se utilizan, pero

lo suficientemente adaptables como para que puedan ser útiles y aplicables a cada caso particular.

Un marco de trabajo diseñado con este propósito es el modelo GE²AN para la gestión transversal de los intangibles empresariales. Desarrollado por Peláez y Casado, se trata de una herramienta que permite comprender como los bienes intangibles de las empresas afectan a las relaciones entre los consumidores y las compañías, tal y como ponen de manifiesto los informes INTED [34]. El modelo GE²AN fue diseñado para la gestión de manera holística de los intangibles de las organizaciones, y permite establecer relaciones causa-efecto en los entornos corporativos en los que existe una fuerte presencia de variables intangibles tales como emociones y sentimientos. Este modelo consta de cinco capas que son: Gestión, Experiencia, Emociones, Actitudes y Negocio que conecta los valores tangibles e intangibles de las compañías [35], [36].

Como se puede apreciar en la figura 3, el modelo GE²AN se diseñó para ser interpretado desde dos perspectivas al mismo tiempo. Por una parte, desde el punto de vista del consumidor, es una abstracción del proceso de decisión del consumidor previamente descrito, en el que los estímulos externos son procesados junto con las experiencias y emociones previas presentes en la memoria del individuo, causando una intención de compra por parte del consumidor. Por otro lado, desde el punto de vista de las compañías, el modelo muestra como las decisiones de negocio que estas llevan a cabo se convierten en experiencias disponibles para sus *stakeholders*, con el objetivo de provocar emociones hacia la empresa y sus productos, que a su vez producen un efecto tangible en las ventas y valores de negocio de la empresa.



Figura 3: Modelo de gestión de intangibles GE²AN [36].

La perspectiva de negocio del modelo GE²AN es especialmente útil para las empresas, ya que provee una manera de cuantificar activos intangibles mediante la relación que estos tienen con los activos tangibles. La primera capa del modelo es la capa de negocio, que engloba toda la información relacionada con las decisiones que se llevan a cabo dentro de la compañía. Las decisiones de negocio son de sobra conocidas por la empresa que las toman y suelen estar disponibles en forma de reportes internos, actas de reuniones y cualquier otro documento que la empresa emplee para registrar su actividad habitual. Lo mismo es aplicable a la última capa del modelo GE²AN, la capa de Negocio, que representa los resultados de la compañía tras todo el proceso, información que está disponible en forma de ingresos, cuentas de beneficios y otros indicadores económicos fruto del desempeño normal de una empresa.

En contraposición a estas dos capas, la experiencia, la emoción y la actitud son los componentes puramente intangibles del modelo GE²AN y juntos conforman el submodelo E²A utilizado para describir el comportamiento de estas variables. Este submodelo presenta varios retos a la hora de medir la información, ya que los activos a los que hace referencia no están disponibles directamente junto con la información habitual de gestión de las empresas y por lo general debe ser adquirida desde fuentes externas como son las encuestas de consumo o los ecosistemas digitales. Además, dado que la información obtenida de esta manera es de naturaleza heterogénea y está generalmente expresada en lenguaje natural, debe ser preprocesada para obtener datos medibles que puedan explicar la relación entre la capa de gestión y la capa de negocio del modelo.

El modelo GE²AN es un marco de trabajo general aplicable a empresas de cualquier sector, por lo que no define ninguna metodología concreta para obtener la información necesaria para medir la experiencia, emoción y actitudes de los públicos, ni tampoco incluye un método para transformar esta información en variables cuantitativas. Una forma de atajar este problema es hacer uso de técnicas demoscópicas como encuestas y procesos de minería de datos para obtener información acerca de experiencias, emociones y actitudes de los consumidores; y con los datos obtenidos proceder a modelar el comportamiento de la componente E²A como un modelo de toma de decisión multicriterio (MCDM).

1.7. La obtención de información acerca de los consumidores

Uno de los primeros pasos necesarios para modelar el proceso de compra de los consumidores es obtener la información de opinión provista por los propios consumidores. Atendiendo a su modo de obtención, existen dos grandes categorías de información: información solicitada e información no solicitada.

La información solicitada es aquella que es preguntada directamente a los públicos mediante encuestas, estudios de mercado, entrevistas y a la que los públicos responden directamente, siendo plenamente conscientes de las preguntas que se les están realizando.

Por otro lado, la información no solicitada es aquella que los usuarios generan de manera natural y voluntaria sin necesidad de estímulos directos por parte de entrevistadores. En esta segunda categoría encontramos la información proveniente de Redes Sociales, la metainformación que se genera al navegar por Internet, información acerca de reservas y pre-ventas de productos o los análisis que los usuarios realizan en diversas páginas web.

Dado que la forma de obtener los datos de opinión provenientes de los consumidores es muy distinta dependiendo de si se trata de información solicitada o no solicitada, en la actualidad la mayoría de los estudios no son capaces de operar con ambas al mismo tiempo y se centran en una de estas dos formas de obtención de información, sin embargo, ambas tienen sus ventajas e inconvenientes.

1.7.1. La obtención de información solicitada

La información solicitada es aquella información que se obtiene de manera directa preguntando a los grupos de interés. La principal ventaja de este tipo de información es que se puede diseñar una campaña de obtención de información exhaustiva acerca de cualquier temática, independientemente de la presencia que tenga un tema concreto en los ecosistemas digitales.

La característica principal de los métodos de obtención de información solicitada es el hecho de que la persona o grupo de personas de las que se quiere obtener información son conscientes, o al menos tienen una idea general, del uso que se le va a dar a la información que proveen en el momento de dar sus respuestas. Esto tiene la ventaja de que la información que proveen

puede focalizarse en los puntos clave que más necesiten conocerse a la hora de realizar el estudio; pero por otro lado cuenta con la problemática adicional del aspecto estratégico de las decisiones, ya que los encuestados, conociendo el uso que se le va a dar a la información que ellos mismos provean, pueden optar por respuestas más en consonancia con sus objetivos o lo que se considere socialmente más aceptable en lugar de responder con su verdadera opinión, especialmente cuando se trate de temas sensibles [37].

Aunque existen varias técnicas para obtener información solicitada, tales como los grupos focales y las entrevistas, el método más común para la obtención de información solicitada es la realización de encuestas. Una encuesta consiste en la elaboración y despliegue de un cuestionario con el propósito de extraer información de una población de personas concreta. Las opiniones de grandes grupos de personas se obtienen preguntando de manera explícita a una muestra representativa de este grupo y extrapolando los resultados a la población completa.

Dependiendo de las preguntas en el cuestionario, la encuesta puede enfocarse en diversas temáticas y extraer información acerca de preferencias, opiniones o comportamientos entre otros datos de interés. Y es por ello por lo que esta técnica de obtención de información es de uso extensivo en diversos campos tales como la educación, la sociología, el marketing o la medicina [38]–[41].

Para crear un cuestionario, es necesario seleccionar correctamente la información que se desea obtener, las técnicas que se van a emplear para contactar con los entrevistados, elaborar una serie de borradores, revisiones y pruebas hasta que se pueda garantizar que las preguntas seleccionadas y su redacción son verdaderamente capaces de extraer la información que se necesita de las personas entrevistadas. A la hora de redactar el cuestionario hay que tener en cuenta diversos factores que puedan afectar a las respuestas dadas por los entrevistados, tales como el orden en el que se le presentarán las preguntas y sus opciones, que su redacción no introduzca sesgos, la utilización de preguntas abiertas o cerradas y la duración de la encuesta entre otros.

Además de la propia elaboración del cuestionario, la creación de la encuesta debe contemplar aspectos como la selección de la muestra, de manera que esta sea representativa y el tipo de encuesta que se va a realizar, así como el método de captura de datos entre múltiples aspectos [42].

Respecto al método de captura, las encuestas pueden ser realizadas de manera presencial por un entrevistador, ya sea mediante lápiz y papel (PAPI), apoyados por dispositivos móviles (CAPI) o de manera telemática ya sea por vía telefónica (CATI) o mediante una encuesta web (CAWI). Las encuestas presenciales, son las que presentan una mayor fiabilidad y tasa de

respuesta, aunque son las más costosas dado que, por lo general, requieren el desplazamiento de personal debidamente entrenado para registrar físicamente las respuestas de los entrevistados. Por otro lado, las encuestas tipo CAWI no suelen requerir un entrevistador, pero la tasa de respuestas suele ser bastante más baja que en los otros casos, lo que debe ser tenido en cuenta a la hora de la elaboración de la muestra [43].

En el caso de ecosistemas digitales, debido a su naturaleza, el tipo de encuesta más habitual suele ser el CAWI. Este tipo de encuestas autoadministradas presentan una serie de retos que pueden afectar a la calidad de la información que recogen. Entre estos retos encontramos, además de la ya mencionada baja tasa de respuestas, un mayor efecto del orden en el que se muestran las opciones al encuestado, la dificultad de obtener una muestra verdaderamente representativa de la población objetivo y respuestas menos elaboradas en el caso de preguntas abiertas, especialmente cuando estas se rellenan desde dispositivos móviles. Es por ello, que la elaboración de las preguntas del cuestionario, su presentación y la selección de la muestra cobran una mayor importancia a la hora de presentar encuestas online, si se desea garantizar el suficiente rigor como para que estas puedan ser consideradas una fiel representación de la opinión real de la sociedad y los consumidores a la hora de tomar decisiones [44].

1.7.2. La obtención de información no solicitada

La información no solicitada es aquella que se obtiene acerca de uno o varios actores sociales sin incitar directamente a los grupos de interés, entendiéndose como actor social, aquella entidad acerca de la cual se generen conversaciones y opiniones de manera espontánea. La forma más habitual de obtener este tipo de información es mediante la captura de información que se genera de manera orgánica en las conversaciones habituales en los ecosistemas digitales acerca de las distintas temáticas. Aunque es posible obtener información no solicitada a través de otros medios, tales como registros de transacciones comerciales, geolocalización, sistemas de vigilancia y diversos tipos de sensores.

Debido a que la información no solicitada comprende un grupo muy heterogéneo de datos, el tipo de información disponible y su uso varía en cada caso. Por ejemplo, la información de negocio tradicional se ha utilizado desde bastante antes de que esta estuviera siquiera computarizada para la realización de auditorías internas que puedan ayudar en la toma de decisiones [45]. Sin embargo, con el auge de los ecosistemas digitales, en los últimos años se ha despertado un creciente interés por la obtención y procesamiento de información proveniente de este medio que puedan ser útiles para las empresas.

Trabajar con datos provenientes de ecosistemas digitales conlleva una serie de ventajas, la más destacable de ellas es que de forma orgánica se genera constantemente un flujo de información que se encuentra disponible de manera pública en Internet. En las Redes Sociales, particularmente, podemos encontrar opiniones acerca de prácticamente cualquier temática, ya que la mayoría de sus usuarios expresan constantemente sus opiniones acerca de productos y servicios.

La mayoría de los consumidores que interactúan con información acerca de productos y servicios escriben comentarios puntuales acerca de sus opiniones, experiencias y emociones, mientras que algunos influenciadores más especializados proveen análisis detallados acerca de las características y capacidades de los mismos [46]. Debido a esto, las fuentes de información no solicitada permiten obtener con facilidad datos que luego deberán ser procesados para obtener información acerca de los activos intangibles del negocio. Una ventaja adicional de este proceso es que, debido a que la información de este tipo ha sido vertida de manera espontánea por los públicos, estos tienden a expresar sus opiniones sin ningún tipo de tapujos acerca de los temas que les resultan de mayor interés. De esta manera, se evita el aspecto estratégico de las decisiones presente a la hora de trabajar con información solicitada [47].

El auge de los datos masivos supuso grandes cambios en la forma de analizar la información. A la hora de trabajar con información solicitada, se recurre al muestreo como medida para paliar la escasez informativa derivada de la interacción necesaria con los entrevistados. Esta muestra es una limitante para la realización de análisis más granulares como por ejemplo a la hora de trabajar con sub-mercados y subcategorías. Por el contrario, cuando se emplean datos masivos provenientes de información no solicitada, la gran cantidad de datos disponible permite dividir la información en diferentes grupos para apreciar detalles más pequeños.

Otro elemento a considerar cuando se trabaja con grandes volúmenes de datos es la exactitud de la información. A menudo, los datos masivos resultan confusos, de calidad variable y están distribuidos en una infinidad de servidores a lo largo del planeta, lo que puede afectar a la calidad de los datos. Estos datos son numerosos en cantidad, pero a cambio suelen ser de menor calidad que los datos aportados por las técnicas demoscópicas empleadas a la hora de trabajar con información solicitada. Aunque existen técnicas que se pueden emplear para medir esta calidad de los datos [48], su uso depende de las necesidades de cada estudio, ya que en la mayoría de las ocasiones lo que se busca es una tendencia general. En esos casos, puede ser deseable asumir la pérdida de exactitud a nivel micro a cambio de una mayor percepción del nivel macro, optando así por una reducción en la calidad de los datos a cambio de una mayor cantidad [49].

Precisamente, es la enorme cantidad de información la que presenta uno de los mayores problemas a la hora de realizar un análisis de información. Prácticamente la mitad de la población mundial goza de acceso a Internet y el tráfico IP mensual se mide actualmente en cientos de exabytes, lo que sumado a la heterogeneidad y calidad de los datos hace inviable el procesamiento de toda la información disponible mediante los procedimientos de análisis de datos tradicionales [33]. La respuesta al problema del tratamiento de cantidades masivas de datos vino en la forma de las técnicas Big Data, diseñadas para el tratamiento de grandes cantidades de datos provenientes de diferentes fuentes.

El objetivo de las técnicas de Big Data es la transformación de datos masivos en información útil que permita alcanzar el conocimiento necesario para que las personas puedan tomar las decisiones correctas. Este proceso se puede apreciar en la figura 4. Los datos son simples observaciones acerca del estado de las cosas, ya sea en forma de números, texto, imagen, etc. son cuantificables y fácilmente capturables por las máquinas. La información se puede definir como datos provistos de relevancia y propósito. Para llegar a obtener información, los datos deben pasar por procesos de análisis que permitan representar la información de manera que pueda ser comprensible para aquellos que la visualizan. El conocimiento es una abstracción de alto nivel que representa información interpretada, en su contexto, provista de un significado por las personas que la interpretan [50].



Figura 4: Transformación del dato e información y conocimiento

El proceso básico para pasar del dato al conocimiento se muestra a continuación en la figura 5. Este proceso presenta múltiples retos, incluyendo inconsistencias de los datos, datos incompletos, escalabilidad, tiempo disponible para el procesamiento de los datos y consideraciones de seguridad. Para poder realizar un análisis los datos deben estar bien contruidos. Sin embargo, considerando la variedad de los datos presentes en los problemas de Big Data, se hace necesario realizar un preprocesamiento de la información utilizando

distintas técnicas, entre las que se incluyen el *data cleaning*, la integración de datos, la transformación de los mismos, la reducción de ruido y la eliminación de inconsistencias. Cada uno de estos subprocesos presenta sus propios retos [51].

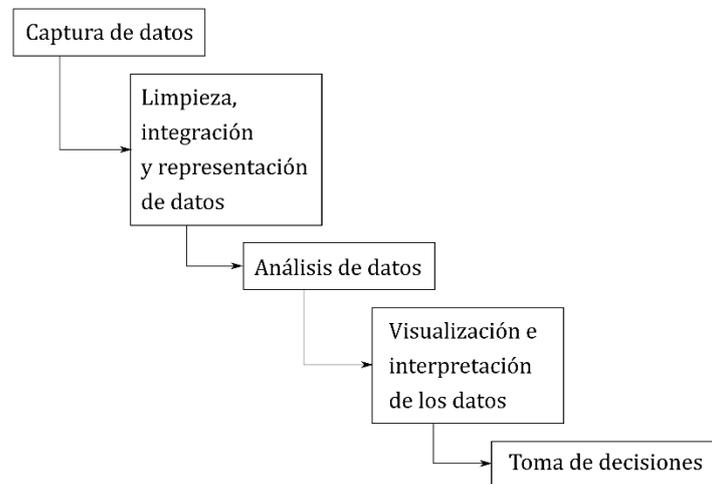


Figura 5: Proceso de descubrimiento del conocimiento

Cuando se trabaja con información no solicitada, el primer problema a atajar en las primeras fases del proceso de descubrimiento de conocimiento es la obtención de datos. Estos datos se pueden extraer de múltiples fuentes tales como bases de datos, ficheros, páginas web, entre otros. Estas fuentes no son homogéneas ya que cada una cuenta con su propia arquitectura y estarán diseñadas para un propósito concreto, por lo que generalmente no estarán optimizadas para el tipo de análisis que pretendemos realizar.

Atendiendo a la morfología de los datos [52], distinguimos tres categorías. Por un lado, encontramos los datos estructurados, que se pueden almacenar por filas y columnas como tablas en una base de datos tradicional, suele ser el tipo de datos más predominante en información de ventas, transacciones, encuestas, etc. En segundo lugar, los datos semiestructurados, que, aunque no se ajustan a un esquema fijo y explícito, suelen disponer de una sintaxis que permite facilitar la interpretación de los mismo. En esta segunda categoría encontramos, entre otros, archivos JSON y XML además de la mayor parte de la información proveniente de dispositivos IoT (*Internet of Things*). Por último, al tratar con información no solicitada es muy habitual utilizar fuentes de datos no estructurados, que son más complejos de procesar ya que se presentan en un formato difícil de manipular para una máquina, son el texto en lenguaje natural, las imágenes o el contenido multimedia ampliamente disponible en Redes Sociales, blogs y foros.

El almacenamiento de grandes cantidades de información presenta sus propios problemas dependiendo de la solución empleada, tales como el coste del soporte de almacenamiento, la velocidad de lectura/escritura de los discos o de la red en el caso de almacenamientos NAS (del inglés, *Network Attached Storage*) o SAN (del inglés, *Storage Area Network*) y la escalabilidad de la solución de almacenamiento. Esto, sumado a la necesidad de asegurar la calidad de los datos, hace que debamos ser selectivos con la información que se almacena y hacerlo de manera que la misma pueda ser lo más útil posible.

Debido a las limitaciones anteriormente descritas, lo más habitual es combinar las dos primeras capas del proceso del descubrimiento de conocimiento mostrado en la figura 5 mediante un proceso de extracción, transformación y carga (ETL) [53] que permita preprocesar y almacenar la información potencialmente útil en almacenes de datos más adecuados para su posterior análisis tal y como se muestra en la figura 6. De esta manera, la capa de extracción puede implementar diversas soluciones para obtener datos provenientes de fuentes muy distintas entre sí, que se llevan a un área de *staging* o preparación donde se realizan diversas transformaciones con estos datos, de manera puedan ser guardados en un almacén de datos en una forma que minimice el impacto causado por la heterogeneidad de las fuentes.

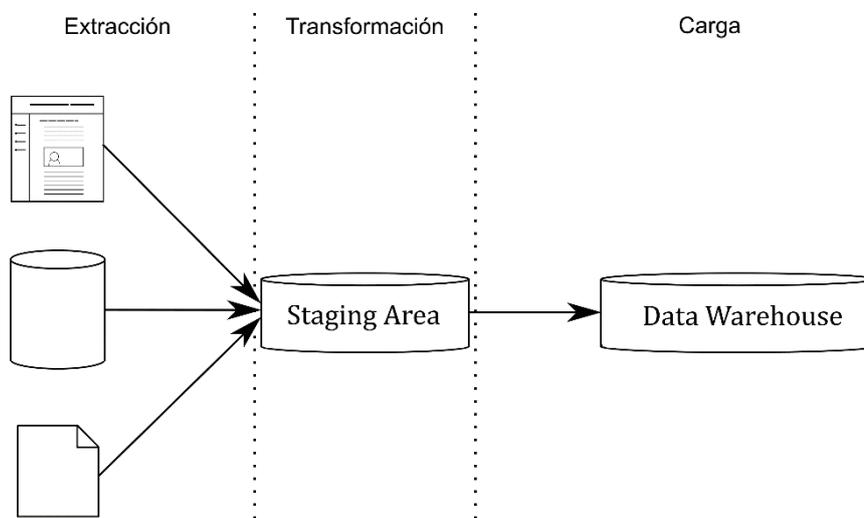


Figura 6: Proceso ETL

En la capa de transformación del proceso ETL se suelen realizar principalmente transformaciones básicas como son la selección de campos de la base de datos, cambios de formato de ciertas variables como fechas y unidades de medida que varían entre regiones del mundo o cálculos de claves de búsqueda y hashes. También en esta etapa es habitual encontrar

operaciones más complejas como son el cálculo de valores derivados, como es el valor de sentimiento de una comunicación, la validación de ciertos tipos de datos en busca de incongruencias o valores perdidos y la combinación de datos provenientes de diversas fuentes entre otros. No toda la información que entra a la capa de transformación es necesariamente almacenada en la base de datos para su posterior análisis ya que se pueden implementar una serie de filtros para descartar la información que pueda resultar poco útil o redundante [54].

En la última fase del proceso ETL, la fase de carga, se almacena la información en el formato que sea más adecuado para su posterior análisis. Igual que la implementación de la fase de extracción es muy dependiente de las fuentes, para la fase de carga es necesario tener en cuenta los análisis que se planean realizar a la información almacenada. Para las aplicaciones de Big Data en tiempo real, aunque en algunos casos es todavía recomendable el uso de bases de datos SQL tradicionales, últimamente suele ser habitual el empleo de esquemas más flexibles que las bases de datos tradicionales; como son las bases de datos NoSQL que en comparación con las primeras permiten, para tareas relacionadas con el Big Data, modelos de datos más flexibles junto a un alto grado de escalabilidad y una mayor velocidad de inserción y consulta [55].

1.8. El análisis de la información de los ecosistemas digitales

Siguiendo con el proceso de descubrimiento del conocimiento anteriormente descrito, una vez se han obtenido los datos y estos han sido preparados para ser procesados, el siguiente paso, tal y como se muestra en la figura 5 es llevar a cabo un análisis que dependerá tanto de los datos obtenidos como del conocimiento que se desea adquirir acerca de los mismos.

En un análisis de Redes Sociales (del inglés *Social Network Analysis* - SNA), se trabaja con información proveniente de los ecosistemas digitales en general y de los social media en particular, con el objetivo de comprender las experiencias, emociones y actitudes de las personas que interactúan en las Redes Sociales digitales. Dicho de otro modo, el SNA es una herramienta que nos permite acceder a las variables intangibles del modelo GE²AN para comprender, o incluso, influir en los públicos. Cuando este análisis se aplica con el objetivo de desentrañar el comportamiento del consumidor, las metas más habituales suelen ser lograr una masa crítica de audiencia a través de medios digitales, fomentar la interacción con la marca, descubrir las temáticas más habituales durante las interacciones o dirigir la intención de compra hacia una serie de productos y servicios.

Para cada uno de los casos anteriores, las métricas más relevantes para el análisis serán distintas; en el caso de la interacción de la marca, lo más habitual es observar el número de comentarios e interacciones con los mismos (*likes, fans, followers, etc.*). Para fomentar la interacción se suele analizar principalmente lo anterior junto con el perfil de las personas que interactúan con la marca y su competencia en busca de influenciadores. A la hora de descubrir las temáticas más habituales lo más común es el ranking de los temas de conversación junto con el análisis de sentimiento de los mismos para conocer si se habla positiva o negativamente acerca de estos temas y por último, a la hora de dirigir la intención de compra se analiza la actividad de la propia marca y su sector, haciendo especial énfasis en el impacto que tienen diversas actividades como son las campañas publicitarias o los lanzamientos de nuevos productos [56].

Con estos datos, son muchos los análisis que se pueden realizar y las métricas que se pueden emplear, pero generalmente los elementos que las empresas desean conocer acerca de sus públicos se encuentran agrupados en tres grandes categorías: el alcance, las temáticas y la intención. El alcance mide la cantidad de interacciones, así como la calidad de los autores de estas, ya que no todas las personas que interactúan ejercen el mismo impacto en los ecosistemas digitales. Las temáticas nos indican acerca de qué hablan los públicos y cómo hablan acerca de estos aspectos incluyendo aspectos como el sentimiento acerca de las mismas. Por último, la intención es un diagnóstico de la relación de estos comportamientos con respecto al valor de negocio que pretende evaluar (ya sea de manera directa o indirecta) las posibles ventas que se realizan del producto; entre los aspectos que se contemplan en esta categoría se incluyen medidas que están estrechamente relacionadas con las mismas como pueden ser las visitas a la tienda, el número de reservas de un producto y la cantidad de nuevas suscripciones a la página del producto.

Como se puede apreciar, la cantidad de información disponible es abrumadora y esto dificulta el poder conocer la apreciación de un bien de manera inmediata. Además, cuando se mide la reputación de una marca, es necesario también comprobar la evolución de esta apreciación por parte de los públicos a lo largo del tiempo si queremos conocer en su contexto los efectos que los procesos de gestión tienen desde el punto de vista de los *stakeholders*. Es por este motivo que existe una creciente demanda de índices de reputación online (del inglés *Online Reputation Index - ORI*) que puedan agregar todas las opiniones en una cantidad mínima de variables, cuya observación continuada permita la rápida comprensión del estado de una entidad social [57].

1.9. Los indicadores de reputación online

Los indicadores de reputación online (ORIs) son medidas proxy de valores intangibles que tratan de determinar el valor de una persona, empresa o marca, a partir de la información de opinión obtenida acerca de los públicos. Actualmente se han convertido en una de las fuentes de prescripción más usadas por los consumidores en las compras en línea para formarse opiniones acerca de productos y servicios.

Sitios web como *Amazon*, *eBay*, *TripAdvisor*, *Trivago* y *Booking*, entre otros, utilizan escalas como son las estrellas, *likes* y valoraciones que permiten a los potenciales consumidores conocer la satisfacción que tienen consumidores previos acerca de los productos y servicios que se encuentran actualmente en oferta, convirtiendo a quienes han adquirido los productos en prescriptores sociales de los mismos [58], [59].

La inmediatez que supone conocer una opinión resumida en un número o una etiqueta, especialmente si agregan múltiples opiniones en un solo valor, es una de las características más apreciadas por los consumidores de este tipo de índices, porque permiten hacerse una idea general del grado de satisfacción que podría proveer cada producto. El éxito de páginas web que agregan rankings de múltiples sitios, como por ejemplo *Metacritic* o *RottenTomatoes*, es una clara prueba de ello [60].

Por contra, el resumir en un solo valor todas las características de un producto conlleva una enorme pérdida de información que en muchos casos puede producir el efecto contrario al deseado por el consumidor, creándole expectativas sobre un producto que no se corresponden con la vara de medir utilizada por un individuo en particular. Los propios sitios web son conscientes de esta limitación.

Algunos ejemplos de cómo estos sitios web han tratado de atajar este problema son: *Metacritic* y *RottenTomatoes*, que separan las críticas de la prensa de las opiniones vertidas por usuarios debido a sus diferentes puntos de vista [61], páginas web como *3DJuegos* o *Eurogamer* que eliminaron directamente el sistema de puntuaciones para centrarse en las recomendaciones mediante texto y plataformas como *Netflix* permiten a sus usuarios otorgar valoraciones positivas y negativas a sus productos, pero los utilizan únicamente para alimentar sus sistemas internos de recomendación sin mostrar resultados agregados [62].

Todas estas dificultades que tienen las plataformas web a la hora de utilizar ORIs no son más que una consecuencia de la naturaleza multidimensional que tienen los actores sociales. Precisamente, cuando se vierten opiniones través de Internet, muchas de las características de

estos actores sociales quedan reflejadas en estas opiniones, pero no así en los ORIs. En ocasiones las opiniones sobre un mismo producto, incluso aunque provengan de un mismo individuo, pueden parecer contradictorias ya que no todas las características de un mismo producto serán igual de atractivas, lo que se traduce en que las críticas no sean completamente binarias, sino que contengan al mismo tiempo opiniones tanto positivas como negativas que dependerán de la característica de la cual se esté hablando [63].

Las compañías que ofertan productos y servicios centran muchos de sus esfuerzos de marketing y publicidad en que el consumidor sea consciente de sus características, especialmente aquellas que resultan en una ventaja competitiva, antes de que realicen la compra, pero también deben prestar especial atención a no exagerar demasiado las expectativas del consumidor. Durante el proceso de decisión de compra, el consumidor tiene en cuenta todos los aspectos que conoce acerca del producto, lo que le genera unas expectativas que influirán en su decisión.

Para un consumidor, el no conocer correctamente las características del producto puede llevar a que sus expectativas sean muy diferentes a la realidad. Si son muy bajas, será difícil que el proceso de decisión de compra se concrete en una compra real. Por el contrario, si las expectativas son muy altas y tras comprar el producto estas no se cumplen, la experiencia del consumidor se verá afectada negativamente en la fase post-compra del proceso. Habida cuenta de que las opiniones están muy fuertemente arraigadas en las experiencias y emociones que se han experimentado y que el comportamiento tras la venta influye en futuras ventas, el no conocer y atajar correctamente estas expectativas puede reducir notablemente la apreciación de una marca.

Sin embargo, como hemos visto anteriormente, las medidas habituales que se emplean a la hora de hablar de ORIs, como son las estrellas, el número de *likes* y otros elementos similares ya han demostrado ser insuficientes para contener adecuadamente la información necesaria para modelar la decisión de compra, hasta el punto de ser eliminadas de algunas plataformas de valoración de contenido. Una posible solución a este problema es la elaboración de indicadores más centrados en atajar cada una de las características de los productos y servicios, de manera que puedan ser útiles para evaluar una familia concreta de productos. Gracias a este enfoque, en la literatura científica podemos encontrar múltiples ORIs de probada eficacia a la hora de comparar fácilmente elementos pertenecientes a una gran variedad de entidades como son universidades, destinos turísticos, productos financieros o la elección del centro hospitalario [64]–[67].

Sin embargo, el tener que utilizar múltiples indicadores diferentes entre sí, plantea un nuevo problema. Los encargados de tomar las decisiones en una empresa necesitan, además de la

información, inmediatez en la interpretación de esta, especialmente en el contexto online donde el periodo de inicio de una crisis reputacional se mide en minutos. Bien utilizados, los ORIs consiguen precisamente eso a través de la elaboración de rankings, pero para ello hay que asegurarse en su elaboración que tienen en cuenta los criterios en los que se basan los consumidores para tomar decisiones y son capaces de abordar realmente las preguntas que las empresas necesitan responder.

1.10. La toma de decisión multicriterio.

En nuestro día a día, las personas realizamos de manera innata procesos de toma de decisiones, comparando múltiples alternativas de acuerdo con diferentes criterios. La mayoría de estas decisiones son intrascendentes y las tomamos sin darnos siquiera cuenta de que estamos llevando a cabo este tipo de procesos, como ocurre con las más de 200 decisiones de media que tomamos a diario relacionadas solamente con la comida [68]. Sin embargo, algunas de las decisiones que tomamos pueden causar grandes impactos tanto en nuestra vida como en la vida de otras personas, por lo que para asegurar que este tipo de decisiones son lo más correctas posibles, se hace necesario conocer en profundidad cómo funciona este proceso interno, cómo nuestra mente toma las decisiones y cómo se pueden modelar de manera formal los problemas de decisión.

La toma de decisión multicriterio es la rama de la investigación operativa que trabaja el modelado de los problemas de toma de decisión. Un problema de decisión, por lo tanto, consiste en la identificación de la opción más apropiada de entre un conjunto de alternativas, atendiendo a una serie de criterios fijados por personas que tienen intereses concretos y conocimiento del objetivo a conseguir. Cuando se trabaja con un grupo de personas, a menudo, la importancia de los criterios es distinta entre distintos decisores, lo que imposibilita encontrar una solución capaz de satisfacer plenamente a todo el mundo, obligándonos a seleccionar la opción que satisface la mayoría de las preferencias del grupo. Especialmente en este tipo de problemas es conveniente el uso de técnicas formales y objetivas para la toma de decisión, de manera que se pueda asegurar que se ha elegido la mejor opción posible de entre las alternativas disponibles.

Existen múltiples métodos en la literatura científica que permiten modelar los problemas de MCDM entre los que destacan el *Analytical Hierarchy Process* (AHP) y su generalización, el *Analytic Network Process* (ANP), la *Simple Multi-Attribute Rating Technique* (SMART), la

teoría de conjuntos difusos, el método para la *ELimination Et Choix Traduisant la REalité* (ELECTRE) y el *Preference Ranking Organization METHod for Enrichment of Evaluations* (PROMETHEE). Todos estos métodos fueron concebidos con el objetivo de identificar las soluciones más deseables para un problema de decisión en situaciones en que los decisores puedan tener intereses que compitan entre sí [69].

La aplicación de modelos de MCDM en el campo de la investigación de mercados es bastante evidente ya que, conociendo el proceso interno de toma de decisiones, estos métodos nos permiten simular el proceso que un consumidor lleva a cabo en el momento que toma la decisión de compra. Por ello, estos métodos se han utilizado en múltiples ocasiones para determinar la importancia que los consumidores dan a los criterios en el momento de evaluar los productos y servicios disponibles en el mercado. A modo de ejemplo, este tipo de métodos se ha empleado satisfactoriamente para evaluar nuevos mercados, elegir proveedores, optimizar flujos de trabajo, reclutar nuevos empleados y medir la satisfacción de los consumidores [70]–[77].

El uso de técnicas MCDM ha permitido modelar procesos de negocio complejos con buenos resultados, y esto ha permitido entender ciertos aspectos del comportamiento del consumidor en diferentes ámbitos. Sin embargo, la adaptación de los métodos MCDM para este tipo de problemas no es sencilla, dado que existen múltiples técnicas de MCDM y es imprescindible decidir cuidadosamente cuál de ellas emplear y cómo adaptarla a cada problema en particular. Además, en la mayoría de los casos hace falta contar con un panel de expertos que extraigan la información de criterios y alternativas para ser evaluados.

Esta necesidad de seleccionar un método y adaptarlo a las particularidades de cada problema es lo que hace que llevar a cabo un proceso formal de toma de decisiones quede relegado a aquellos problemas importantes que pueden causar un efecto a largo plazo en múltiples personas. Sin embargo, aunque sería deseable que existiese un método universal para la toma de decisiones multicriterio que pudiese ser aplicado al conjunto de todos los problemas de decisión, la existencia de un modelo de estas características es imposible debido al teorema de la imposibilidad de Arrow.

1.10.1. Teorema de la imposibilidad de Arrow

Para entender el teorema de la imposibilidad de Arrow, lo primero es partir de la base de que el objetivo de cualquier proceso de toma de decisión multicriterio es la selección de la opción más deseable de entre un grupo de posibles alternativas. Además, cuando existen múltiples

decisores, el proceso de toma de decisión tiene la responsabilidad añadida de asegurar que la decisión colectiva es justa para todos los participantes que intervienen en el proceso. Por este mismo motivo, Arrow define un conjunto de cinco condiciones, o axiomas, que se tienen que cumplir si queremos que el ranking resultante de aplicar un método MCDM refleje las opiniones de todos los decisores [78].

- Eficiencia de Pareto: Si una opción es preferida por todos los decisores, esta preferencia deberá mantenerse en el orden final. Esto quiere decir que, dadas dos opciones cualesquiera A y B , si para todo decisor se cumple que $A \succ B$, entonces en el ranking final se tiene que cumplir que $A \succ B$.
- No dictatorialidad: No puede haber un dictador en el proceso de toma de decisiones. Esto quiere decir que no hay ningún decisor particular que disfrute de una posición dominante en cuanto a que sus preferencias deban ser preservadas por encima de las preferencias expresadas por los demás decisores.
- Dominio no restringido: Los decisores no tienen ninguna restricción en el orden individual de las alternativas. El resultado final del ranking debería ser único y determinista.
- Independencia de alternativas irrelevantes: Introducir o eliminar una alternativa no debería producir una reversión de rango en las otras opciones. Esto es, dado un ranking en la forma $\{A, B\}$ al añadir una nueva opción X , el ranking resultante deberá preservar la preferencia existente $A \succ B$. Asimismo, dado un ranking $\{A, B, C\}$, la eliminación de la opción B también deberá preservar la preferencia de las opciones restantes, en este caso $A \succ C$.
- Transitividad: El orden del ranking final debe ser consistente con respecto a las preferencias, por lo que si $A \succ B$ y $B \succ C$, entonces debe darse que $A \succ C$.

Aunque todos los axiomas anteriores deberían satisfacerse por cualquier método MCDM que asegure equidad en el proceso de decisión, el teorema de la imposibilidad de Arrow demuestra matemáticamente que es imposible satisfacer todas estas condiciones al mismo tiempo [79]. A consecuencia de esta demostración, no es posible crear un método de toma de decisión universal que sea completamente justo para todos los posibles problemas de decisión, lo que hace que la decisión más acertada dependa del método que hemos escogido para llegar a ella. Esto nos lleva a una de las mayores críticas de los modelos MCDM: Incluso con las mismas entradas, distintos métodos pueden producir rankings diferentes y potencialmente conflictivos.

Algunos trabajos previos han tratado de atajar este problema, derivado del teorema de la imposibilidad de Arrow, mediante el uso de múltiples técnicas de agregación para combinar los resultados de múltiples métodos de MCDM en un ranking único [80], sin embargo, la

creación de un super-ranking que aglutine los resultados de múltiples MCDM no deja de ser un nuevo modelo MCDM sujeto también al teorema de la imposibilidad de Arrow, por lo que la opción más extendida a la hora de modelar la toma de decisión multicriterio es relajar algunos de los axiomas de Arrow para obtener una solución que sea lo suficientemente buena para el subconjunto de problemas concretos que se pretenda atajar.

En nuestro caso, el objetivo es determinar las preferencias de los consumidores, mediante un único método, a través del modelado del comportamiento del consumidor durante la fase de evaluación de alternativas del proceso de decisión de compra. Esto presenta un reto adicional, ya que los métodos actualmente disponibles fueron diseñados para utilizarse en entornos controlados en los que los decisores que participan están disponibles para consultas y aclaraciones acerca de sus preferencias. Sin embargo, en este trabajo se plantea el uso de información no solicitada vertida en ecosistemas digitales, lo que inhabilita el uso de la mayoría de los métodos existentes para este mismo propósito. Por lo tanto, es necesario desarrollar modelos más específicos de MCDM diseñados expresamente para comprender las preferencias de los consumidores, que también puedan atajar problemas específicos a la toma de decisiones dentro del marco de las transacciones comerciales que se exponen a continuación.

En primer lugar, aunque es posible obtener información solicitada proveniente de los consumidores preguntándoles acerca de los criterios que utilizan para realizar sus compras, no es realista obtener información compleja que requiera un conocimiento previo acerca de los modelos de decisión multicriterio, como son la influencia de las sinergias que se dan entre los distintos criterios y que también influyen en la intención de compra. En segundo lugar, aunque es habitual que se realicen sondeos a los consumidores para conocer su satisfacción en la fase post-compra, habitualmente se utilizan ORIs como estrellas y escalas de Likert que pueden producir resultados indeseados como reversiones de rango. Por último, dado que la información no solicitada en forma de comentarios representa la mayoría de la información proveniente de Internet, es necesario disponer de métodos que sean capaces de obtener de manera rápida, automática y determinista, las preferencias implícitas en estos comentarios para poder alcanzar un conocimiento realmente útil acerca de lo que opinan los consumidores de los productos.

1.10.2. Métodos de MCDM

Como ya se ha adelantado, existen diversos métodos de toma de decisión multicriterio en la literatura científica que son aplicables a múltiples áreas. El progreso tecnológico de las últimas décadas ha incrementado el interés en disponer de técnicas que puedan ser utilizadas para

comprender el proceso de toma de decisiones de las personas y en cómo este se podría modelar de una forma automatizable, lo que causó la creación y adopción de distintos métodos de MCDM como los anteriormente mencionados AHP, ANP, SMART, teoría de conjuntos difusos, ELECTRE, PROMETHEE entre otros. Debido a la gran cantidad de métodos disponibles y a que la elección de estos métodos depende del contexto en el que se va a aplicar, centraremos nuestra atención en aquellos que han demostrado ser particularmente útiles en el estudio del comportamiento del consumidor [81].

1.10.3. El método AHP

El *Analytic Hierarchy Process* (AHP) y su generalización, el *Analytic Network Process* (ANP) conforman una familia de teorías matemáticas que permiten obtener vectores de prioridad a partir de matrices recíprocas positivas, que representan valores de comparaciones entre pares realizadas por los decisores. Estos métodos utilizan los autovectores para obtener los vectores de prioridad y así presentar un ranking de alternativas [82].

Para obtener un ranking, se construye una jerarquía de criterios (en el caso de AHP) o una red de criterios (en el caso de ANP) y posteriormente se asigna un peso relativo a cada uno de los criterios dado por la importancia que los decisores otorgan a cada uno de los criterios en comparación con los demás. Habitualmente, la forma de hacerlo es utilizar un cuestionario diseñado exprofeso para extraer los valores necesarios para construir las matrices de comparación entre pares.

Una matriz de comparación entre pares A representa las relaciones de preferencia entre pares de criterios o alternativas siguiendo una escala, donde $X = \{1, 2, \dots, n\}$ es el conjunto de n criterios o alternativas a ser comparados. La matriz $A_{n \times n}$ de comparación entre pares tendrá entonces la siguiente forma:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad (1)$$

Donde a_{ij} expresa la preferencia de x_i con respecto a x_j valorado en una escala. La escala más utilizada con diferencia es la definida por Saaty y mostrada en la tabla 1, junto con las inversas multiplicativas de todos los valores que aparecen en dicha tabla, los cuales representan una

relación invertida, por ejemplo, $a_{ij} = 1/3$ significa que j es ligeramente más preferido que i [83].

VALOR DE a_{ij}	SIGNIFICADO
1	i son j igualmente preferidos.
2	valor intermedio entre 1 y 3.
3	i es ligeramente más preferido que j .
4	valor intermedio entre 3 y 5.
5	i es más preferido que j .
6	valor intermedio entre 5 y 7.
7	i es bastante más preferido que j .
8	valor intermedio entre 7 y 9.
9	i es absolutamente más preferido que j .

Tabla 1: Valores de la escala de Saaty con sus significados

Dado que las preferencias entre criterios o alternativas son recíprocas, cualquier matriz de comparación entre pares A deberá preservar también esta reciprocidad por lo que para cada par de i y j , debe cumplirse que a_{ij} sea la inversa de a_{ji} con respecto a la escala elegida. Esto también provoca que el único elemento posible en la diagonal principal de la matriz A sea la identidad. Aunque generalmente los métodos de la familia de AHP definen una relación multiplicativa entre los elementos de la escala, también existen variantes que han conseguido modelar satisfactoriamente problemas con relaciones aditivas, de conjuntos difusos o intervalares [84]–[86].

El proceso de rellenar la matriz de comparación entre pares pasa por ir preguntando al decisor acerca de sus preferencias entre todas las parejas posibles de criterios, y una vez disponemos de una matriz de comparación entre pares A completamente rellena, ya es posible obtener el autovector principal de la matriz A . Este autovector, una vez normalizado, resultará en el ranking de los criterios pertenecientes a X .

La complejidad de rellenar esta matriz de comparación entre pares se incrementa de acuerdo con el número de criterios que se consideren para el problema de decisión. Para un conjunto de n criterios el número de comparaciones total a realizar será de $(n^2 - n)/2$ por cada decisor que vayamos a incluir en el proceso. Además, para problemas complejos AHP y ANP establece agrupaciones de criterios en distintos niveles de sub-criterios, cada uno con sus

propias matrices de comparación entre pares, por lo que en ocasiones no es realista pretender rellenar todos los valores de todas las matrices de comparaciones entre pares que se van a emplear en el proceso de toma de decisión. Para estos casos, existen diferentes métodos de reconstrucción que pueden ser utilizados para reducir el número de comparaciones o rellenar matrices incompletas mientras se mantiene un cierto nivel de consistencia [87], [88].

La consistencia de los juicios provistos por los decisores también es una propiedad deseable en un proceso de toma de decisiones. Para AHP y ANP, la consistencia se deriva de la transitividad de los elementos de la matriz de comparación entre pares. Más concretamente, para que una matriz A sea consistente, esta debe satisfacer que $a_{ij} \cdot a_{jk} = a_{ik}$ para $i, j, k \in [1..n]$.

Sin embargo, un grado de consistencia absoluta como el que aquí se plantea es difícilmente obtenible en la práctica. Cuando tratamos con seres humanos con cierto grado de incertidumbre, es casi seguro que aparezcan ligeras inconsistencias en los juicios, especialmente en las matrices más complejas. Por lo que al igual que los métodos de toma de decisión multicriterio tratan de satisfacer el mayor número de requisitos, sabiendo por el teorema de Arrow que no serán capaces de satisfacer el 100% en todos los casos, los índices de consistencia existentes en la literatura científica tratan de establecer el grado de inconsistencia asumible en un proceso de toma de decisiones con personas.

El método para calcular la consistencia de matrices de AHP más utilizado es el CI, definido por el propio Saaty y que hace uso del autovalor principal de la matriz, pero existen otros métodos que permiten un mayor grado de control, disponen de otras propiedades matemáticas deseables o que mejoran la exactitud para algunos tipos de matrices, con lo que una vez se ha decidido utilizar el método AHP para establecer un ranking de criterios, también resulta conveniente seleccionar el método más adecuado para realizar los cálculos de consistencia [89]–[92].

1.10.4. El Eigenvector Method for Pairwise Voting (EMPV)

La dificultad de aplicación del método AHP escala cuando se aumentan de manera excesiva tanto el número de decisores como el número de alternativas. Por lo que en los casos en los que se dispone de un número de decisores muy elevado, puede ser preferible el empleo de otro método de la misma familia que permita simplificar el proceso.

El *Eigenvector Method for Pairwise Voting* (EMPV) es un método basado en la idea de votación entre pares, que considera el orden de las preferencias obtenido a partir de múltiples

decisores para un conjunto concreto de alternativas. La principal diferencia con el método AHP es que en EMPV se construye una única matriz a partir de los rankings obtenidos por cada uno de los decisores, en lugar de una matriz por cada decisor. Para ello, este método define una serie de alternativas $A = \{a_1, a_2, \dots, a_n\}$ y un conjunto de votantes $N = \{1, 2, \dots\}$ que emiten su voto entre cada par de alternativas de acuerdo con sus preferencias.

En este caso, en lugar de emplear una escala para obtener los valores de preferencia, los decisores simplemente indicarán para cada una de las parejas de a_i y a_j el sentido de su preferencia. Una forma de obtener de manera rápida esta información es pedir a cada uno de los decisores que ordene según sus preferencias todas las alternativas incluidas en A .

A continuación, se construye una matriz de votación como sigue:

$$A(\phi) = \begin{bmatrix} 1 & \frac{v_{12}(\phi)}{v_{21}(\phi)} & \dots & \frac{v_{1n}(\phi)}{v_{n1}(\phi)} \\ \frac{v_{21}(\phi)}{v_{12}(\phi)} & 1 & \dots & \frac{v_{2n}(\phi)}{v_{n2}(\phi)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{v_{n1}(\phi)}{v_{1n}(\phi)} & \frac{v_{n2}(\phi)}{v_{2n}(\phi)} & \dots & 1 \end{bmatrix} \quad (2)$$

Como en los otros métodos de la familia de AHP, el autovector asociado al autovalor máximo de la matriz $A(\phi)$ es el vector que representa de la manera más precisa la preferencia conjunta de los diferentes votantes.

Al contrario que AHP, que evalúa pares de criterios utilizando una escala, en EMPV los decisores pueden emitir directamente un ranking de alternativas completo sin la ayuda de una escala, y los rankings de todos los decisores pueden ser utilizados para construir una única matriz que contenga la información provista por cada uno de los decisores. Dado que la tarea de los decisores se ha reducido a simplemente indicar el sentido de sus preferencias, se sacrifica la precisión cuando hay un número pequeño de decisores, pero se incrementa la facilidad con la que estos pueden ser incorporados al proceso de decisión, lo que lo convierte en un buen método para ser aplicado a los ecosistemas digitales, que por naturaleza tienden a contar con grandes cantidades de usuarios que disponen de una cantidad limitada de tiempo.

1.10.5. Teoría de conjuntos difusos

La lógica difusa es una extensión de la lógica clásica diseñada para resolver problemas relacionado con información imprecisa y datos inciertos. Los seres humanos, por naturaleza, estamos acostumbrados a trabajar con información incompleta, y la toma de decisiones es precisamente la tarea más habitual que debemos realizar partiendo de conocimientos inexactos. En la mayoría de las ocasiones, resulta imposible disponer de toda la información para poder tomar una decisión. Obtener toda la información existente acerca del tema del que se pretende tomar una decisión, en no pocas ocasiones puede requerir un coste en esfuerzo mucho mayor que el beneficio esperado, cuando no se trata de una tarea directamente imposible.

Por lo tanto, y teniendo en cuenta que la mayoría de los problemas de decisión multicriterio tienen que lidiar con esta falta de información, tiene sentido que la teoría de conjuntos difusos haya demostrado ser una herramienta valiosa para tratar problemas de decisión complejos, obteniendo resultados muy consistentes en problemas en los que intervienen varios decisores con objetivos distintos y en los que hay que llegar a una decisión que satisfaga a la mayoría de los implicados. Sin embargo, los métodos difusos tienden a ser algo más difíciles de implementar como herramientas computacionales aplicables a los problemas de la vida real [81].

Un problema de toma de decisión multicriterio en el campo de la lógica difusa se puede representar como un conjunto de n alternativas $X = \{x_1, x_2, \dots, x_n\}$ y otro conjunto de m expertos $E = \{e_1, e_2, \dots, e_m\}$, donde cada experto expresa su relación de preferencia entre elementos pertenecientes a X , de manera que $p_k \subset X \times X$, y $\mu_{p_k}(x_i, x_j) \in [0,1]$ indican el grado de preferencia de la alternativa x_i sobre la alternativa x_j .

Normalmente, los expertos no son capaces de expresar los grados de preferencia de manera natural como un valor numérico real, y dado que estamos trabajando con valores imprecisos tampoco tendría sentido que pudiese hacerlo de esta forma. En estos casos, se suele proveer un conjunto de etiquetas lingüísticas de manera que estos sean capaces de expresar su preferencia de una manera más parecida a cómo funciona el lenguaje natural.

Existen muchas formas de crear el conjunto de etiquetas lingüísticas y asignarle un valor, teniendo en cuenta la granularidad que los distintos expertos sean capaces de expresar y las características del problema. Lo más habitual es emplear funciones que permitan transformar las etiquetas seleccionadas por los expertos en lenguaje natural a números difusos que representen un rango de pertenencia. Dependiendo de las capacidades y el conocimiento previo de los expertos, es posible utilizar distintos niveles de granularidad en las etiquetas para

el proceso de valoración que estos realizan, complicando ligeramente el proceso de transformación hacia números difusos, pero facilitando el proceso de valuación a los expertos [93].

Una etiqueta lingüística puede representarse de tres formas distintas: un conjunto difuso, un intervalo o como un único valor numérico, siendo la forma más extendida la primera de estas tres opciones mediante el uso de números difusos [94]. Un número difuso representa un conjunto de posibles valores con un peso asociado en el intervalo $[0, 1]$ indicando el grado de certeza con el que cada valor numérico continuo pertenece al conjunto.

En el caso de los problemas de toma de decisión multicriterio que utilizan números difusos, las elecciones más comunes de números difusos son los números difusos triangulares (TFN), siendo utilizados, por su sencillez computacional, en cerca del 90% de los artículos en la literatura científica [95].

Los TFN permiten definir las etiquetas lingüísticas mediante el uso de ternas, en la forma (t_1, t_2, t_3) donde $t_1 < t_2 < t_3$, que representan intervalos de pertenencia a una clase, siendo t_1 y t_3 los extremos de la clase y t_2 el punto central [96]. En la tabla 2 se muestra un ejemplo de un TFN con cinco etiquetas lingüísticas que podrían utilizarse para obtener información de preferencias a partir de un experto mediante una pregunta del tipo “¿Con qué grado considera que la alternativa a_i es mejor que la alternativa a_j ?”

ETIQUETA LINGÜÍSTICA	TFN
NADA	(0.00, 0.00, 0.00)
POCO	(0.00, 0.23, 0.43)
MEDIO	(0.30, 0.50, 0.70)
MUCHO	(0.57, 0.77, 1.00)
COMPLETAMENTE	(1.00, 1.00, 1.00)

Tabla 2: Ejemplo de aplicación de TFN con 5 etiquetas lingüísticas

Los pesos de las cinco etiquetas lingüísticas definidas en la tabla anterior se pueden representar gráficamente como se muestra en la figura 7.

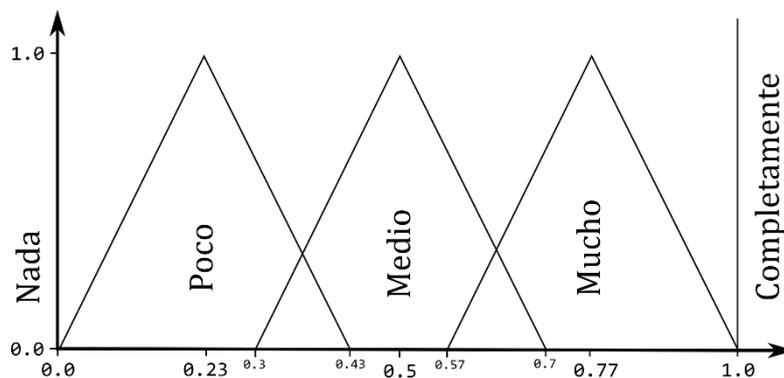


Figura 7: Representación gráfica de los valores lingüísticos definidos en la tabla 2

Una vez se ha establecido una función de pertenencia con la forma $\mu_R: X \times X \rightarrow [0, 1]$ indicando el grado de preferencia entre dos alternativas, necesitamos agregar la información obtenida a partir de los expertos pertenecientes al conjunto E para obtener una decisión en grupo. Para este propósito, es necesario elegir una función de agregación que pueda garantizar que la decisión tomada tenga en cuenta, de manera justa, las valoraciones emitidas por los expertos. Al expresar el problema en términos de lógica difusa e introducir valores de intensidad en las preferencias, evitamos forzar a los decisores a emitir valores fijos, lo que introduce un mayor grado de libertad a la hora de seleccionar un modelo de agregación que cumpla lo máximo posible con los axiomas de Arrow cuando utilizamos operadores *Ordered Weighted Averaging* (OWA) [97].

1.10.6. Operadores OWA

Los operadores *Ordered Weighted Averaging* (OWA) son una familia de operadores que permiten la agregación de diferentes elementos dependiendo de la posición ordenada de sus atributos. Dentro de esta extensa familia, encontramos muchos de los métodos de agregación más utilizados, como son el máximo, el mínimo, la mediana y las medias ponderadas. Desde su concepción, los OWA han sido ampliamente utilizados en muchos ámbitos entre los que se incluyen la toma de decisiones, el marketing, el modelado de redes y bases de datos, la inteligencia artificial y una multitud de sistemas que utilizan como motor la lógica difusa [98], [99].

De acuerdo con la definición original dada por Yager [100], un operador OWA de dimensión n es una función $F: I^n \rightarrow I$ donde $I = [0, 1]$ que tiene un vector W de pesos asociados

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (3)$$

tal que $w_i \in [0, 1]$ y $\sum_i w_i = 1$ y además

$$F(a_1, a_2, \dots, a_n) = \sum_j w_j b_j \quad (4)$$

siendo b_j el j -ésimo mayor valor perteneciente a la colección a_1, a_2, \dots, a_n . Considerando un n -vector B de órdenes en el que cada elemento $b \in [0, 1]$ y $b_i \geq b_j$ si $i > j$. Dado un operador OWA con un vector de pesos W y una n -tupla de argumentos (a_1, a_2, \dots, a_n) , podemos asociar esta tupla con el vector B de órdenes de manera que B sea el vector de los argumentos de F en orden descendente. Utilizando esta notación obtenemos que:

$$F(a_1, a_2, \dots, a_n) = W'B \quad (5)$$

De manera que los pesos de cada elemento de W quedan asociados a una determinada posición en lugar de a un valor particular, dado que w_i es el peso asociado al i -ésimo mayor elemento.

En la práctica, esto se traduce en que según los valores dados a W se pueden implementar múltiples operadores de agregación y estos respetarán características de linealidad, monotonía, simetría e idempotencia que caracterizan a los operadores OWA y todas ellas deseables para su utilización como operadores de agregación aplicables a métodos de MCDM [100].

Como consecuencia, son muchas las semánticas que pueden asociarse a los operadores OWA, siendo su uso más habitual la creación de operadores que puedan ser utilizados como medidas de tendencia, dado que estos generalmente requieren que se satisfagan las condiciones de idempotencia, monotonía y conmutatividad, como es el caso de los operadores de mayoría, tales como el MA-OWA o el SMA-OWA [101], [102]. Sin embargo, otros trabajos han descrito los pesos de los operadores OWA como un tipo especial de distribución de probabilidad, asumiendo $A = \{a_1, \dots, a_n\}$ como un conjunto de posibles resultados e introduciendo una distribución de probabilidad P que indique la probabilidad del mejor i -ésimo resultado [103]. La interpretación de los operadores OWA como una distribución de

probabilidad también permite extrapolarlos al campo de la toma de decisión multicriterio, reflejando las actitudes de los decisores hacia las alternativas y criterios en el proceso decisión.

1.11. Cómo comprender la decisión del consumidor

Cuando un consumidor finalmente decide la adquisición de un bien o servicio, son muchos los factores que intervienen. A la hora de modelar este comportamiento como un proceso de MCDM, los diferentes productos competidores al alcance del consumidor son modelados como alternativas, y estos factores que afectan a la decisión son los criterios del modelo de toma de decisión multicriterio elegido. Sin embargo, determinar y valorar los criterios vinculados al proceso de decisión de compra es una tarea compleja.

Incluso aunque los estímulos que influyen a los consumidores puedan ser observados a partir de las interacciones en los ecosistemas digitales, una gran parte de los estímulos son internos y están imbuidos en la memoria del consumidor, siendo afectados por sus conocimientos y experiencias, lo que dificulta esta medición. Los distintos canales disponibles para obtener la información presentan un reto adicional, ya que para poder comprender qué criterios tienen realmente en cuenta los consumidores durante la elección de un producto o servicio, es necesario estudiar las opiniones de ellos en todas sus formas posibles, independientemente de si esta aparece en forma de opinión solicitada u opinión no solicitada.

1.11.1. La combinación de información solicitada y no solicitada

La mayor ventaja de la información solicitada es que los sujetos pueden ser consultados directamente acerca de las áreas de interés del estudio de manera que, si se utilizan los métodos correctos, obtenemos información estructurada y mucho más precisa que la que se obtiene a partir de la información no solicitada. Sin embargo, existen ciertos problemas que encontramos al emplear información solicitada. En primer lugar, este tipo de información es generalmente más difícil de obtener, por lo que las muestras obtenidas tienden a ser mucho menores que cuando se trabaja con información no solicitada. En segundo lugar, la elaboración de cuestionarios es una tarea que requiere personal experto en la materia a consultar que también posea conocimientos estadísticos y psicológicos. De no ser así son muchos los sesgos que se

pueden introducir derivados de errores de muestreo, elaboración de las preguntas y otros elementos que pueden influir indirectamente en las respuestas de los encuestados. Por último, relacionado con lo anterior, encontramos el aspecto estratégico de las decisiones, que consiste en que, para cierto tipo de preguntas, la persona encuestada tiende a responder de la manera que le parezca más favorable hacia su persona o aquello que se perciba socialmente como la respuesta “correcta”, en lugar de dar una opinión desprovista de sesgo [104]–[106]. Todos estos aspectos, de no ser tratados adecuadamente pueden reducir la validez de los datos, por lo que es necesario tomar medidas para evitarlos durante la elaboración de la encuesta.

Por otra parte, la información no solicitada proviene de conversaciones en lenguaje natural, generalmente a partir de comentarios vertidos en Internet sin que se les hayan formulado a los interlocutores preguntas específicas, por lo que se evita el aspecto estratégico de las decisiones, se reducen mucho ciertos tipos de sesgos y es posible obtener muestras lo suficientemente grandes para mitigar al máximo los errores de muestreo o incluso trabajar con poblaciones completas. Sin embargo, en estos casos, el investigador está relevado a un papel pasivo, ya que no interviene y se limita a observar el comportamiento de los consumidores en su entorno natural, por lo que las preguntas respondidas pueden no ser las que el investigador realmente desea conocer.

Lo ideal sería utilizar una aproximación híbrida en la que se puedan combinar la información solicitada con la no solicitada, de manera que ambos tipos de datos puedan reforzarse mutuamente y se pueda obtener un punto de vista mucho más global de lo que piensa el consumidor a la hora de realizar una compra. Aunque la combinación de ambos tipos de información no sería suficiente para obtener una imagen general y completa acerca del consumidor, ya que faltarían atajar otra serie de problemas como son las sinergias entre criterios de los consumidores, el uso de información combinada supondría un gran paso a la hora de obtener información de calidad.

1.11.2. Las sinergias entre criterios

Una vez que se dispone de información, el siguiente paso es utilizarla para extraer aquellas características o criterios en los que se basan los consumidores para decidirse entre varios productos. Existen múltiples trabajos en la literatura científica que emplean técnicas de MCDM, para extraer información de la importancia de los criterios en distintas categorías de productos y servicios. Estos trabajos, que suelen estar diseñados exprofeso para una industria concreta, parten principalmente de información solicitada y la procesan con la ayuda de un panel de expertos, para identificar cuáles son los criterios que emplean los consumidores y qué

importancia le asignan internamente los mismos a estos criterios a la hora de elegir entre un producto u otro [72], [76], [107]–[109].

Sin embargo, los trabajos publicados hasta el momento consideran a cada uno de los criterios como características independientes de los productos, ignorando las sinergias entre las mismas que sí están presentes en la mente del consumidor a la hora de evaluar los productos. Estas interacciones entre criterios a la hora de ser evaluados en no pocas ocasiones son el factor de elección determinante entre varios productos, por lo que deben ser consideradas a la hora de modelar el comportamiento del consumidor. Por ejemplo, un cliente debe elegir entre cuatro productos (Productos A, B, C y D) que son evaluados en una escala [0, 10] con respecto a tres criterios (diseño, calidad y ahorro) tal y como se muestran en la figura 8.

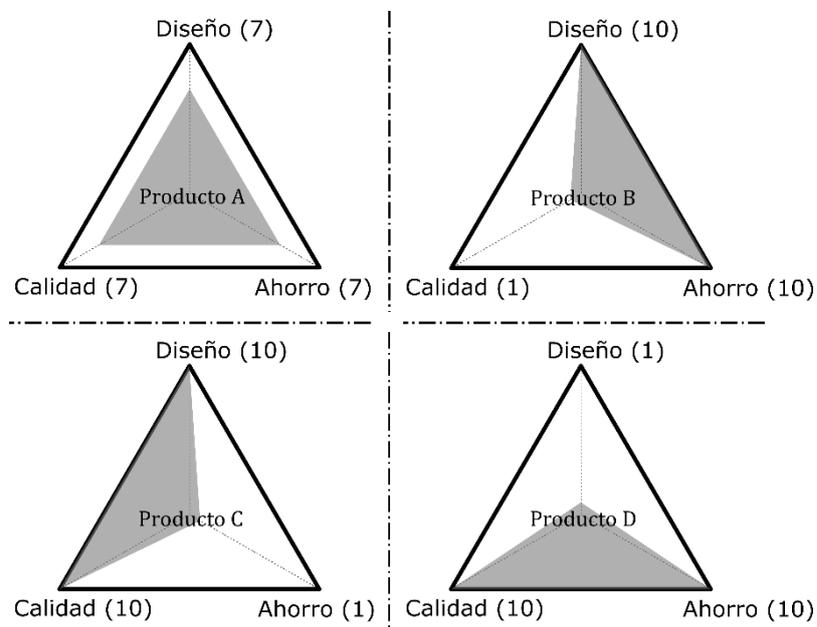


Figura 8: Valoración de cuatro productos con respecto a “diseño”, “calidad” y “ahorro”

Por supuesto, a la hora de elegir uno de estos productos habría que tener en cuenta otra serie de factores, como son las experiencias y emociones previas del cliente, así como la motivación y el uso que se le piensa dar al producto, pero para simplificar el problema supongamos que, para este cliente concreto, todos los criterios tienen la misma importancia. En ese caso, cabría pensar que dado que en cada uno de los cuatro productos presentes sus valoraciones suman exactamente 21 sobre 30, todas las alternativas debieran ser igual de atractivas.

Sin embargo, el producto A resulta más atractivo que el resto debido a las sinergias presentes entre los criterios de diseño, ahorro y calidad escogidos para este ejemplo. Esta sinergia hace

precisamente que sea más deseable obtener una combinación adecuada de “bueno, bonito y barato” que descartar una de ellas en beneficio de las otras dos.

A la hora de modelar este tipo de problemas como un MCDM, cuando no hay relación entre los criterios, se considera la importancia entre distintos grupos de criterios como aditiva. Esto quiere decir que cuando se toma una decisión entre dos criterios cualesquiera c_1 y c_2 junto con los pesos $w(c_1)$ y $w(c_2)$ que se han extraído para resolver el problema de decisión, el peso de los criterios combinados viene dado por:

$$w(c_1, c_2) = w(c_1) + w(c_2) \quad (6)$$

Sin embargo, esta ecuación solo es aplicable cuando se consideran los criterios como entidades independientes. Cuando los consumidores piensan en las alternativas, los criterios que utilizan para evaluarlas están interrelacionados entre sí, por lo que realmente existe un valor $d \in \mathbb{R}$ que afecta a la interacción entre los diferentes criterios, lo que conduce a que el verdadero peso de los criterios combinados sea:

$$w(c_1, c_2) = w(c_1) + w(c_2) + d \quad (7)$$

Dependiendo del valor de d tenemos tres posibilidades:

- Si $d = 0$, en este caso, los criterios son independientes y se trata simplemente de una relación aditiva entre ambos criterios.
- Si $d > 0$, los criterios c_1 y c_2 incrementan su peso al estar juntos. Estamos tratando con criterios que son más deseables en conjunto que por separado, de manera que se refuerzan mutuamente, como ocurre en el ejemplo anterior con calidad y ahorro.
- Por último, si $d < 0$, los criterios c_1 y c_2 se solapan entre sí y por lo tanto pierden interés cuando estos se dan al mismo tiempo. Un ejemplo es que, cuando una marca es bien conocida y valorada por el público entre otras cosas es porque queda implícito un cierto nivel de calidad, por lo que se daría una mayor importancia cuando estos dos criterios no se dan simultáneamente que cuando si lo hacen.

Aunque este fenómeno ha sido estudiado en la literatura científica, la única solución disponible hasta el momento pasa por extraer información acerca de grupos de decisores expertos [110]–[114]. Este enfoque, aunque es válido para su uso en problemas de toma de decisión en

pequeños grupos cerrados y con información solicitada, es insuficiente para modelar el comportamiento de terceras partes a partir de opiniones no solicitadas fruto de un proceso de *Data Mining*. Para estos casos es necesario una metodología que nos permita extraer directamente las sinergias de criterios a partir de grandes conjuntos de comunicaciones, escritas por consumidores reales que opinen acerca de los productos y servicios, y que nos permitan aproximar de una manera más realista el valor que estos otorgan a los productos y servicios que adquieren.

1.11.3. El valor en las comparaciones de los consumidores

Un aspecto que se repite en todos los modelos que pretenden explicar el comportamiento del consumidor a la hora de tomar una decisión de compra es la valoración de las posibles alternativas utilizando una escala. Es un paso lógico, ya que la valoración de los productos por parte del consumidor es una de las cinco fases de las transacciones comerciales. Sin embargo, este paso es relativamente complejo ya que el valor de un producto o servicio es una característica subjetiva que cada individuo asigna de manera diferente. Además, durante esta asignación de valor entran en juego aspectos psicológicos únicos de cada individuo, lo que puede dar lugar a decisiones que, a priori, pueden parecer irracionales [115].

En la actualidad, la herramienta más utilizada para aproximar el valor de un producto en base a información proveniente de ecosistemas digitales son los ORIs que ya explicamos anteriormente. Aunque los ORIs no indican el valor del producto, estos indicadores pueden ser utilizados como medida proxy del mismo, lo que permite hacerse una idea general del valor que tiene un producto. Un bien con una valoración muy baja tiene una alta probabilidad de ser un mal producto, y otro con una valoración muy elevada posiblemente sea una buena solución al problema que pretende solventar.

Para conocer información de un producto en solitario, los ORIs tradicionales funcionan bien. Pero, los productos difícilmente existen como entes aislados, sino que conviven en un mercado donde compiten múltiples productos competidores con otras soluciones alternativas, por lo que el valor percibido de un producto no se puede tampoco medir como un valor aislado, sino que depende de las comparaciones que los consumidores realizan entre las distintas alternativas.

Este es el motivo de que los ORIs que se emplean actualmente como son las estrellas, puntuaciones y el número de *likes* no sean tan útiles como se desearía para medir el valor y algunas plataformas digitales lleguen hasta el punto de dejar de mostrarlos. Realmente, son

dos los grandes problemas que merman la utilidad de los ORIs tradicionales como sistemas de valoración: a) La diferencia de la escala interna entre diferentes usuarios y b) La disociación entre los ORIs y el concepto de comparación. Por un lado, se está agregando información de múltiples decisores en un solo valor, cuando internamente cada uno de ellos utiliza una escala de valoración distinta. Es decir, a la hora de emitir una valoración las “★★★★☆” otorgadas por dos usuarios distintos rara vez tienen el mismo significado para ambos.

La diferencia de escala interna entre personas es mitigable cuando se utiliza información solicitada de manera exclusiva, ya que las personas que van a proveer esta información pueden ser informadas previamente del funcionamiento de las escalas durante todo el proceso. Para ello es necesaria una cuidadosa redacción de las preguntas y que se les instruya correctamente en el funcionamiento de la escala de valoración utilizada, forzando así a todos los entrevistados a utilizar una escala de valoración interna relativamente similar. Desgraciadamente, cuando se emplea información no solicitada, no existe ningún método que nos permita forzar este uso de una escala interna similar para todos los proveedores de información.

El segundo problema, la disociación entre los ORIs y el concepto de comparación, es independiente del tipo de información que se emplee y está más relacionado con el funcionamiento interno de los ORIs y su uso como medida proxy del valor de un producto a la hora de tomar decisiones. Cuando alguien observa la calificación que se otorga a un producto en un sitio web, lo que realmente desea obtener es información acerca del valor que tiene ese producto para el resto de las personas. Este planteamiento es incorrecto porque el valor que tiene un bien no es una característica aislada del objeto, sino el resultado de una serie de comparaciones de ese mismo bien contra todos los productos similares.

Esta utilización de los ORIs como medida proxy del valor es un error al que estamos muy habituados. Pero, incluso aunque se trate del mismo tipo de objeto, pongamos un teléfono móvil, los ORIs tradicionales no sirven para comparar entre distintos productos, ya que una puntuación de “★★★★☆” en un teléfono de 900€ no implica que este sea peor que otro teléfono de 150€ y puntuado con “★★★★★”. Medidas como las estrellas pierden su significado al intentar transformarlas en un esquema de preferencias, ya que realmente lo que otorgan es una medida de un producto de manera aislada.

Esto dista mucho de la realidad ya que el valor de un producto depende en gran medida de las alternativas existentes al mismo, es decir, se debe medir el valor dentro de un conjunto de alternativas comparables entre sí. El proceso de cálculo de un ORI diseñado para aproximar el valor que los públicos otorgan a un producto debería ser análogo al comportamiento del ser

humano. A la hora de valorar cualquier entidad, lo que realmente hace una persona es realizar una serie de comparaciones con otras entidades similares que conozca y le puedan servir de base para construirse una idea de cómo deben ser las cosas y poder establecer esta valoración. Dicho de otra forma, el valor de un bien depende directamente de la comparación con otros bienes de su misma clase [116].

Si realmente deseamos modelar el comportamiento del consumidor a partir de información proveniente de los ecosistemas digitales, es imprescindible establecer un indicador basado en esta comparación de conjuntos de alternativas comparables y que pueda ser utilizado como medida del valor de un producto. Para ello es necesario construir un ORI que funcione a partir de las comparaciones del producto en su contexto. Siendo este contexto el conjunto de alternativas comparables entre sí en el que convive el producto.

1.12. Objetivos y estructura de la memoria

Hasta este punto se han planteado las características del comportamiento del consumidor en la era digital, explicando cómo se obtiene la información para su modelado y planteando la problemática que tienen los métodos actuales para conseguir modelar este comportamiento correctamente.

Algunos de los problemas que quedan por resolver son los siguientes: En primer lugar, no existe actualmente en la literatura científica ninguna metodología que permita combinar información solicitada con información no solicitada de manera que el resultado se refuerce mutuamente. En segundo lugar, aunque el efecto de las sinergias de criterios afecta a la decisión de compra del consumidor, no existe ninguna metodología que permita extraer directamente el valor de estas sinergias y aplicarlas a la elaboración de rankings, lo que afecta a la precisión de los rankings existentes. Y por último lugar, los ORIs existentes están siendo utilizados como medida proxy del valor percibido de los productos para los consumidores, sin embargo, la construcción de estos ORIs difiere notablemente del proceso que los públicos llevan a cabo al asignar valor a distintos bienes y servicios en su contexto, lo que lleva a errores a la hora de extraer información de valor apreciado por los públicos. Los siguientes capítulos de esta memoria se centrarán en la resolución de estos problemas.

En el segundo capítulo se trata el tema de la combinación de información solicitada con no solicitada. Para ello, se ha creado una nueva metodología capaz de enriquecer la información expresada en escalas de Likert, de uso tradicional en las encuestas, a partir de datos resultantes

de procesos de MCDM que pueden ser utilizados con información proveniente de ecosistemas digitales, tales como el resultado de un EMPV. Esta nueva metodología ha sido probada con datos reales del sector turístico y ha demostrado su capacidad para mejorar los resultados a la hora de elaborar rankings.

El tercer capítulo presenta una solución al problema que plantean las sinergias de criterios en la fase de evaluación de alternativas del proceso de decisión de compra. Para ello se ha creado un modelo basado en la teoría de conjuntos difusos que permite no solo considerar estas sinergias de criterios a la hora de elaborar rankings de compra a partir de información de los consumidores, sino también obtener la información acerca de estas sinergias a partir de información no solicitada de manera automática y sin necesidad de recurrir al uso de expertos. Este modelo también ha sido probado con datos reales, demostrando una mejoría con respecto a otros métodos que no tienen en cuenta las sinergias de criterios.

En el cuarto capítulo se elabora un nuevo Indicador de Reputación Online (ORI) diseñado para determinar la preferencia relativa entre distintas alternativas a partir de las comparaciones que se realizan entre ellas de manera natural en los ecosistemas digitales. Este ORI, basado en agregaciones geométricas de matrices de comparación, provee una medida más realista del valor percibido por los consumidores de un producto en su contexto, al tener en cuenta el conjunto de alternativas comparables en las que convive este producto.

Finalmente, en el último capítulo, se exponen las conclusiones que derivan de esta memoria, así como un resumen de los principales aportes y las líneas de trabajo futuras abiertas por los avances realizados en la misma.

Capítulo 2: La combinación de información

A la hora de combinar datos provenientes de fuentes de información solicitada con fuentes de información no solicitada, uno de los principales problemas es la distinta morfología de la información obtenida. La información solicitada tiende a ser fácilmente discretizable, en parte gracias al uso de escalas cerradas como son las escalas de Likert. La información no solicitada, por el contrario, presenta un formato completamente diferente, debido a que suelen involucrar en su obtención el uso de ORIs, técnicas MCDM y otra serie de técnicas que no están diseñadas para su uso con la demoscopia tradicional. En este trabajo se ha presentado una nueva metodología que permite combinar ambos tipos de datos gracias a la inclusión de información de rankings obtenidos mediante EMPV en cuestionarios tradicionales realizados con escalas de Likert. Esta estrategia mixta permite extraer patrones de preferencias de los públicos, corregir problemas asociados a la transitividad entre preguntas en cuestionarios y mejorar el análisis de preferencias considerando los axiomas de Arrow, todo ello manteniendo la compatibilidad con los sistemas de encuestas tradicionales.

Article: Pairwise Voting to Rank Touristic Destinations Based on Preference Valuation [117].

Journal: Sustainability

Year: 2019

Pages: 11(21), 5873

DOI: 10.3390/su11215873

Contributor Roles: Conceptualization, formal analysis, investigation, methodology, software, visualization, writing – original draft.

Capítulo 3: Detección y cómputo automático de sinergias de criterios

En este capítulo, se presenta una metodología novedosa basada en una combinación de técnicas de toma de decisión multicriterio pertenecientes a la familia AHP, herramientas provistas por la teoría de lógica difusa y el uso de la integral de Choquet como método de agregación. El resultado es una metodología completa que permite obtener los pesos de los criterios de manera automática a partir de información no solicitada, detectando y cuantificando, además, el efecto de las sinergias que se producen entre los criterios que utilizan los públicos para evaluar los productos y servicios que se encuentran a su alcance. Esta metodología ha sido probada con información real obtenida directamente de los ecosistemas digitales, demostrando que, al considerar las sinergias de criterios, causa una apreciable mejoría en la predicción de los rankings de compra.

Article: Estimating the importance of consumer purchasing criteria in digital ecosystems [118].

Journal: Knowledge-Based Systems

Year: 2018

Pages: 252-264

DOI: 10.1016/j.knosys.2018.07.023

Contributor Roles: Conceptualization, formal analysis, investigation, software, visualization, writing – original draft.

Capítulo 4: Índice matricial geométrico de comparación: Un ORI construido sobre la comparación de alternativas

En este capítulo se presenta un nuevo índice de reputación online (ORI), basado en agregaciones geométricas de matrices de comparaciones (CMGI), que es capaz de extraer de manera directa información de preferencias a partir de comparaciones expresadas en lenguaje natural. La metodología propuesta es parametrizable, lo que permite su uso con cualquier conjunto de alternativas comparables entre sí. Además, todo el proceso es determinista y parametrizable, facilitando su despliegue como herramienta para monitorizar los cambios en tiempo real del mercado para la pronta detección de crisis reputacionales. Aunque el CMGI ha sido diseñado con los ecosistemas digitales en mente, este ORI está preparado para su uso con información procedente de cualquier tipo de fuente susceptible de albergar información comparativa, independientemente de si esta es solicitada, no solicitada o una combinación de ambas.

Article: Comparison matrix geometric index: A qualitative online reputation metric [119].

Journal: Applied Soft Computing

Year: 2020

Pages: 106687

DOI: 10.1016/j.asoc.2020.106687

Contributor Roles: Conceptualization, formal analysis, investigation, methodology, software, visualization, writing – original draft.

Capítulo 5: Conclusiones y líneas futuras

Como se ha detallado en el estudio de la situación actual al principio de este trabajo, el análisis del comportamiento del consumidor es una ciencia compleja y multidisciplinar que ha experimentado muchos cambios en los últimos años debido a la evolución de la relación empresa-cliente. Para comprender la extensión de estos cambios y aprovechar el máximo posible la información disponible, es necesario que existan modelos que permitan conocer y computar el efecto que producen los ecosistemas digitales en el consumidor.

En este trabajo se han analizado los problemas inherentes al modelado del comportamiento del consumidor a partir de información proveniente de ecosistemas digitales, detectando las debilidades existentes en la metodología empleada actualmente y reforzándolas mediante el aporte de métodos complementarios que puedan solucionar estos problemas de una manera más eficiente y precisa. Tras el estudio del estado actual se han detectado tres grandes problemas que dificultaban la obtención de conocimiento acerca del comportamiento del consumidor, todos ellos apreciables durante la fase de evaluación de alternativas que este lleva a cabo cuando realiza una compra.

Por una parte, la información que se estaba empleando hasta ahora era incompleta ya que, debido a las diferencias en las características de los datos aportados por distintas fuentes, no existía ninguna forma de combinar adecuadamente la información solicitada con la información no solicitada. En el capítulo 2 de esta memoria, se ha establecido una metodología para combinar información en forma de ranking, tal y como se obtiene a través de los análisis de ecosistemas digitales, con datos obtenidos a partir de técnicas demoscópicas tradicionales, gracias al empleo del modelo de votación entre pares EMPV. El uso combinado de esta información complementaria ha demostrado una mejora apreciable en la calidad del conocimiento que se puede obtener acerca de las decisiones que toman los consumidores.

Por otro lado, a la hora de elaborar estos rankings, las interacciones entre los criterios que los consumidores emplean a la hora de evaluar los productos presentaban un reto sin resolver que, hasta ahora, había dificultado el conocer el motivo por el cual los consumidores eligen ciertos productos respecto a otros. Tras una exhaustiva revisión de la literatura científica, se puede apreciar cómo ninguna metodología era capaz de extraer estas sinergias que ocurren de manera natural, cuantificarlas y aprovecharlas para mejorar la predicción de la intención de compra.

La nueva técnica de cálculo de sinergias de criterios aportada en el capítulo 3 de este trabajo extrae los pesos de los criterios y las alternativas directamente de un set de comunicaciones y las utiliza para construir una medida difusa que represente las diferentes combinaciones de pesos, sin asumir asociatividad, que posteriormente son agregados mediante el uso de la integral de Choquet. El realizar el proceso de esta forma permite, de manera automática, y en base a las relaciones implícitas en las valoraciones emitidas por los públicos, cuantificar el efecto de las sinergias de criterios y utilizarlas para elaborar rankings de alternativas de compra.

Con la inclusión de los datos sobre sinergias, la nueva metodología aportada ha demostrado no solo que la presencia de sinergias de criterios en las decisiones de compra afecta al ranking final de alternativas, sino que, el considerar el efecto de estas sinergias acerca aún más los resultados de los rankings predictivos de compra a los valores de ventas reales de los productos cuando estos gozan de una cierta presencia en los ecosistemas digitales.

Por último, se ha detectado que el uso más extendido de los ORIs tradicionales, es decir, como medida proxy del valor apreciado de un producto por los públicos, acarrea una serie de problemas derivados tanto de las diferentes escalas internas que utilizan las personas para valorar, como de la asociación del concepto de “valor” de un producto con alternativas comparables. Esto provoca que medidas como las estrellas o los *likes* sobre un producto no sirvan para comparar entre productos, de manera que, aunque gracias a la metodología aportada en el capítulo 3 podamos extraer información sobre el peso de los criterios a la hora de tomar una decisión, el valor percibido de un producto se vea ofuscado por esta interpretación de los ORIs tradicionales como medida proxy.

Para poder obtener información más realista sobre el valor apreciado que tienen los productos dentro de su mercado, se ha desarrollado un nuevo ORI, presentado en el capítulo 4, que está construido desde cero basándose en el funcionamiento de los procesos psicológicos que los consumidores emplean para asignar valor a los productos. Esto es, a partir de comparaciones entre grupos de alternativas comparables. El CMGI permite extraer esta información directamente a partir de opiniones expresadas en lenguaje natural, tal y como lo harían los públicos, y siguiendo una metodología completamente automatizable y determinista, elaborar un ranking que refleja la apreciación relativa que tienen los públicos para cada una de las alternativas disponibles.

Aunque estas tres metodologías pueden ser utilizados de forma estanca, su verdadero potencial aparece cuando se utilizan de manera combinada, tanto entre ellas como con otras técnicas existentes para el tratamiento de la información, para modelar la fase de evaluación de alternativas, presente en todo proceso de decisión de compra. Mediante el uso conjunto de estas tres técnicas se pueden extraer los criterios que los consumidores emplean, incluyendo

las sinergias implícitas en los mismos y conocer el valor que los públicos otorgan a los productos disponibles, todo ello empleando la máxima cantidad de información posible gracias a una combinación de información solicitada y no solicitada.

Uno de los mayores problemas dados por el rápido crecimiento de los ecosistemas digitales es la ingente cantidad de información débilmente estructurada disponible, presentándose como una amalgama de datos que deben ser procesados por empresas e instituciones si realmente quieren conocer lo que demandan los públicos de ellos. El conocer los criterios y alternativas en los que estos públicos se basan, para obtener decisiones de consumo, es un paso importante para poner algo de orden en todo este caos y será una pieza clave para un futuro cada vez más comunicado.

Una de las líneas de investigación más prometedoras de las que se derivan de este trabajo es precisamente la integración de estos métodos en los marcos de trabajo que ya existen, tales como el modelo GE2AN, así como su combinación con las técnicas demoscópicas tradicionales que ya se vienen utilizando. Como se expuso en el capítulo de introducción, son muchos de los trabajos en el campo de la investigación de mercados que están focalizados a un tipo concreto de producto, por lo que puede ser interesante la elaboración de nuevos modelos específicos aprovechando las capacidades de los métodos de combinación de información, cálculo de sinergias y obtención de valoración subjetiva junto con otros modelos que consideran además aquellas características que consideran el tipo de producto concreto.

Además, aunque en este trabajo nos hemos centrado en el proceso de decisión de compra, los métodos desarrollados son también aplicables fuera del ámbito del estudio del comportamiento del consumidor. Tanto la metodología para la detección de sinergias aportada y el CMGI, como ORI basado en el concepto de comparabilidad pueden tener múltiples aplicaciones en otras áreas fuera del estudio del consumidor tales como la ingeniería industrial, la medicina o la planificación urbanística, entre otras. En estas áreas son muchos los criterios que hay que tener en cuenta a la hora de tomar decisiones que en muchos casos son irreversibles, por lo que merece la pena estudiar la aplicabilidad de los métodos aquí presentados en otras ramas de la ciencia.

Chapter 5: Conclusions and future lines

As we detailed at the beginning of this work, the analysis of consumer behavior is a complex and multidisciplinary science that has undergone many changes in recent years due to the evolution of the company-customer relationship. In order to understand the extent of these changes and to take advantage of the information available to us as much as possible, new computational models are needed to understand and evaluate the effect of digital ecosystems on consumers.

In this work we have analyzed the problems of modeling consumer behavior based on information from digital ecosystems, detecting existing weaknesses in the methodology currently used and reinforcing them by providing complementary methods that can solve these problems in a more efficient and accurate way. After studying the current state of the art, three major problems have been detected that make it difficult to obtain knowledge about consumer behavior, all of them appreciable during the evaluation phase of alternatives that the consumer carries out when making a purchase.

On the one hand, the information that was being used up to now was incomplete, due to the differences in the characteristics of the data provided by different sources. There was no way of adequately combining the solicited information with unsolicited information. In chapter 2 of this work, a methodology has been established to combine both types of information obtained through the analysis of digital ecosystems, with data obtained from traditional demoscopic techniques. The combined use of this complementary information has demonstrated an appreciable improvement in the quality of the knowledge that can be obtained about the decisions made by consumers.

On the other hand, when generating these rankings, the interactions between the criteria that consumers internally use when evaluating products presented another unresolved challenge that, until now, had made it difficult to understand why consumers choose certain products over others. After an exhaustive review of the scientific literature, it can be seen that no methodology was able to extract these naturally occurring synergies between criteria, quantify them and use them to improve purchase intent prediction. The new technique for calculating criteria synergies provided in chapter 3 of this work allows, in an automatable way, and based

on the implicit relationships in valuations expressed by the public, to quantify the effect of criteria synergies and use them to build rankings of purchasable alternatives.

With the inclusion of data on synergies, the new methodology provided has shown not only that the presence of criteria synergies in purchase decisions affects the final ranking of alternatives, but also that considering the effect of these synergies brings the results of predictive purchase rankings even closer to the real sales values of products when they enjoy a certain level of presence in digital ecosystems.

Finally, it has been detected that the most widespread use of traditional ORIs, i.e., as a proxy measure of the appreciated value of a product by audiences, entailed a series of problems derived both from the different internal scales used by people to value, and the dissociation of the concept of "value" of a product with the valuation of comparable alternatives. This means that measures such as stars or likes on a product are hardly ever useful for comparing between products. Although thanks to the methodology provided in chapter 3 we can extract information on the weight of criteria when making a decision, the perceived value of a product is obfuscated by this interpretation of traditional ORIs as a proxy measure.

In order to obtain more realistic information about the perceived value of products within their market, a new ORI has been developed, presented in chapter 4, which is built from scratch based on the functioning of the psychological processes that consumers use to assign value to products. That is, based on comparisons between groups of comparable alternatives. The CMGI allows to extract this information directly from opinions expressed in natural language, as the public would do, and following a fully automatable and deterministic methodology, to produce a ranking that reflects the relative appreciation that the public has for each of the available alternatives.

Although these three methodologies can be used in a stand-alone fashion, their true potential appears when they are used in combination with each other, or with other existing techniques for information processing, to model the evaluation of alternatives phase, which is present in every purchase decision process. By using these three techniques together, it is possible to extract the criteria that consumers use, including the synergies implicit in them, and to know the value that the public places on the available products, all this using the maximum amount of information possible thanks to a combination of solicited and unsolicited information.

One of the biggest problems caused by the rapid growth of digital ecosystems is the huge amount of weakly structured information available, presenting itself as an amalgam of data that must be processed by companies and institutions if they really want to know what audiences demand from them. Knowing the criteria and alternatives on which these audiences

base their consumption decisions is an important step to bring some order to all this chaos and will be a key piece for an increasingly communicated future.

One of the most promising lines of research derived from this work is precisely the integration of these methods into existing frameworks, such as the GE²AN model, as well as their combination with traditional demoscopic techniques already in use. As explained in the introductory chapter, many of the works in the field of marketing research are focused on a specific type of product, so it may also be interesting to develop new specific models focused on those fields, while taking advantage of the strengths of the presented methods for combining information, calculating synergies, and obtaining subjective valuation.

Furthermore, although in this work we have focused on the purchase decision process, the methods developed are also applicable outside the scope of the study of consumer behavior. Both the methodology for the detection of synergies and the CMGI could have multiple applications in areas not closely related to consumer behavior such as industrial engineering, medicine, or urban planning, since in these areas there are many criteria to be taken into account when making decisions, which in many cases turn to be irreversible.

Apéndice A: English summary

The discovery, creation, and management of demand for products and services is one of the main focuses of any thriving business. Whether it being for launching a new business, a new product or in their day-to-day operation, it has always been important for enterprises to understand the behavior of consumers. Customers want solutions for their needs, therefore, one of the goals of market research is to find or create a need that their target customers have, so companies can offer them goods or services that can be perceived as useful towards the satisfaction of said need. From the point of view of the customer, there are many reasons to make a purchase such as fixing a problem, easing a monotonous task, as an investment or to achieve a social status among others. Some of these reasons are more obvious than others, but if customers don't feel that a need can be satisfied by a product, the product is perceived as useless to them, so they would hardly buy it.

In this regard, information is the basis for a successful business, since it allows provides an understanding not only of the needs of the existing and prospective consumers, but it also serves as window in which we can observe the competition, the current offers available to consumers within a market and their response to past and present products available to them. This information can be used to evaluate management decisions and change corporate behavior to provide experiences in accordance with what consumers perceive as more attractive products.

There are many applications of marketing research techniques on corporate management, such as product research, advertisement, branding research, concept testing, pricing research and exploring consumer satisfaction. Each of these applications bring great value to the companies which invest resources on them. As an example, advertisement can create or enhance this perception of a consumer need and associating it to a product, which makes it an important pillar on the budget of marketing activities for most companies. Advertisement has been proven to have a very positive impact on the intangible value of enterprises and it is widely used to increase the value of brands [10], [11].

To get the maximum out of marketing techniques such as advertisement, knowledge of the potential and target customers is required. More specifically, we must understand the consumer decision process, that is, how and why consumers choose the specific products and services they are going to buy. As described in the scientific literature, this process is comprised of five stages which are understood at an abstract level from the business

perspective. However, to create a truly useful model capable of assessing the real value of intangible assets in an automated way by using computational techniques, we must understand how the consumers' minds work when acquiring goods and services, how we can assess their preferences, attitudes and motivations towards markets, brands and products, what are the sources from which this information can be obtained, which are the limitations that have yet to be addressed and how to overcome them.

Moreover, the internet has undeniably changed the interactions between companies and consumers with economies of leading countries rapidly evolving towards digital ecosystems. This change has affected core aspects of society such as education, science, innovations and, of course, the relationships between companies and customers, by making the world a more interconnected we made each aspect of society more intertwined.

For instance, under these new conditions, partnership of scientists, higher schools and external service providers have become a necessity for many industrial sectors and knowledge about the market is now an essential aspect in every stage of product development, instead of some kind of preliminary work that some companies did only at the beginning stages of product design. The accessibility of information stimulates innovative activity and reduces costs, making the possibility of increasing profits in a growing market very appealing.

This effect works the other way around as in a widely connected society consumers want to know about enterprises as much as enterprises want to know about them. There is a genuine concern from customers on how enterprises operate, how they treat their employees, how they manage environmental issues derived from their activity and so on. In this regard, transparency has become a very important topic among customers, and in some sectors, it has been proved to directly impact profits [13], [14].

Since information has become a main economic asset, we are shifting towards an economy based on knowledge, where this knowledge has become a main factor of the production chain [18]. As such, enterprises are allocating many resources to understanding consumers' behavior in the digital era, how their prospective clients make choose which goods and services they are going to buy and how enterprises should manage their intangible assets.

A.1. Consumer behavior and choice

Consumer behavior is usually modelled as a cognitive process, in which the consumer receives a series of inputs in the form of environmental stimuli competing for the attention of the consumers. With this information, a decision process is carried out inside the consumer's mind, in which all this information is mentally processed along with subjective information about past experiences and emotions that they have stored inside their memories. Beliefs and goals and other intrinsic aspects of consumers vary from one individual to another, causing different behaviors in different customers even when they are exposed to the same stimuli. The output of this process is the consumer attitude towards the product which ultimately translates into the purchase decision [22].

A.1.1. The five stages of commercial transactions

The consumer decision process consists of the 5 stages shown in figure A1, which are: Recognition of a need, search of information, evaluation of product options, purchase decision and post-purchase behavior [23]. These five steps are the basis of the reasoning that is subconsciously made in the customers' mind and are fundamental for the understanding of consumer intentions both before and after purchase and creating brand loyalty.

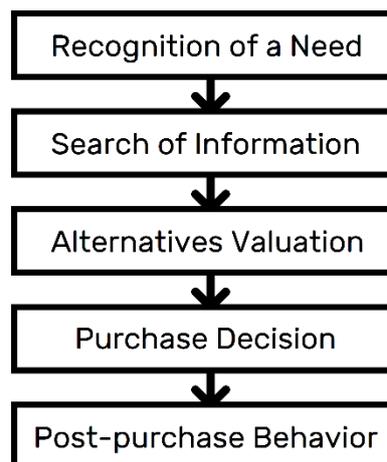


Figure A1: The purchasing decision process

The prospective consumer starts the purchasing process with the recognition of a need. This step is usually performed subconsciously and can include several reasons depending on the particular product to be purchased, which can range from very practical reasons such as hunger, or some need for transportation to very subjective ones such as admiration from their peers or a sense of achievement. Since this step jumpstarts the whole purchasing process, there is a great interest in the scientific literature on how to identify or create such needs, and the different techniques employed through history in order to successfully shifting perceptions of this perceived needs [24]–[26].

Once the consumer has identified a need that has to be satisfied, the second step of the purchasing process is a search of information. This search can be passive, when consumers limit themselves to just being perceptive to the information being provided to them from different sources, or active when they are proactive in the search of a solution for their need by purposefully looking for products and services that might be able to cover their need. As for the nature of the inputs received, they include both objective information such as technical specifications of a product and subjective information, such as opinions on the product from other consumers.

A primary source of opinions on products has always been word-of-mouth, a concept that is defined as any direct communication between a receiver and a communicator that the receiver perceives as non-commercial towards a brand, product or service. Word-of-mouth and has been widely studied since the 1950s, but with the raise of the digital current digital ecosystems such as social media and online reviews, word-of-mouth has widened its scope. While some decades ago consumers had access to just a few dozens of peers with opinions on popular products or brands, the Internet has given access to each consumer to thousands of reviews for even the most obscure products [27].

When used to their advantage, this universal access to information presents many opportunities for businesses. Firstly, consumers are not restricted to what is available near them, but more importantly, they can obtain the information from other consumers about products even if there is nobody geographically near with experiences about said products. And secondly, since online information on reviews and opinions is usually public, it provides businesses with information on how consumers view their products, which can be used to understand how consumers perceive those available currently in the market or which are the current trends in consumer habits [28].

The third step of the consumer decision process is the alternatives valuation phase. In this stage, the consumer makes use of the information obtained in the previous stage to determine which of the available products is the most suitable one to tackle their need. In order to do so,

the prospective consumer evaluates the available alternatives on different aspects according to their memories of previous experiences, beliefs, attitudes, goals, purchasing power, etc. As shown in figure A2, this step can be modelled as a multi-criteria decision-making process (MCDM) by modeling as criteria the aspects of the product that consumers use to evaluate them, resulting in a ranking that carries on to the following step, the purchase decision.

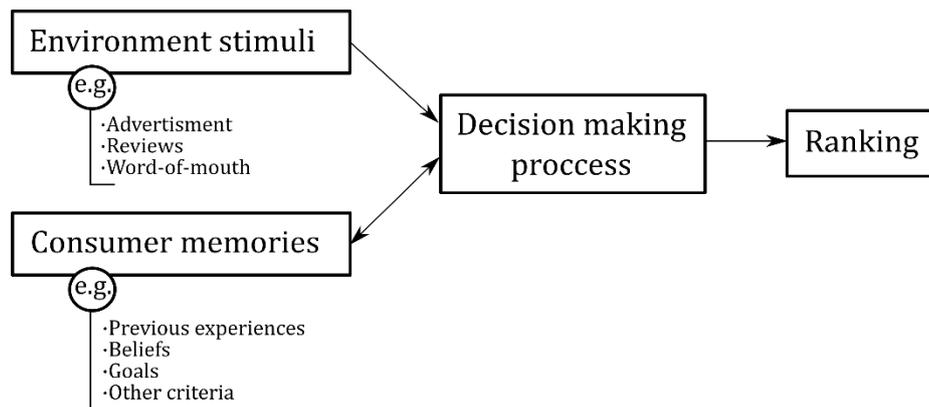


Figure A2: Inputs and outputs of the decision-making process performed on the alternatives valuation step

In the fourth step, the purchase decision stage, the consumer performs an action in accordance with their ranking of preferences. There are multiple actions that the consumer can take, and external factors can still affect the final action the consumer such as the availability of each alternative or other impeding and facilitating conditions. Furthermore, the possible response from the consumer is not binary, aside from purchasing or not purchasing, consumers can choose to postpone the decision, re-evaluate their options and rent or borrow a product among many other options.

The last step of the purchasing decision process is the post-purchase behavior. This step is key because of its implications for future purchase patterns. When a purchase decision is made, it generates new emotions within the consumer, these emotions are retained in the memory of consumers and will become a previous experience for any future purchase. Moreover, emotions related to a previous purchase can be easily transmitted, so any other prospective consumers can pick up this information in the information search phase of their purchasing decision process, affecting the outcome of future purchases.

A.1.2. Intangible assets management: the GE²AN model

There are two categories of assets that can be found in any company. Tangible assets are physical assets that can be owned by a company and their value can be easily quantifiable, computer equipment, machinery and buildings owned by the company are examples of tangible assets. On the other hand, intangible assets don't physically exist and although they have monetary which can also be quantified, these are more difficult to identify and measure. Brand loyalty and product reputation are examples of intangible assets that provide revenue to companies, despite not being physical assets.

In recent years, the importance of intangible assets within companies has been steadily growing, business need to know the perception that their public have about them to adjust their strategies [33]. This is a consequence of enterprises realizing that intangible assets management are key to provide differential value in a worldwide market in which consumers have access to many competitive products. Good intangible management has been proved invaluable in some aspects for companies such as cost reductions, investors attraction, high exchange rates in stock markets, increased customer loyalty, employee attraction and retention and minimizes the impact of crises [31], [32].

For this reason, companies are very interested in building information systems capable of providing a better understanding on their intangible assets for supporting their decision process. These systems should be capable of extracting knowledge from large pools of information obtained from different heterogeneous sources, both internal and external, and process this information accordingly in order to provide a greater understanding of their business model and its relationship with their customers. The GE²AN framework provides a model for the transversal management of intangible assets of enterprises and is a very useful tool to understand how intangible assets affect the relationships between consumers and businesses.

The GE²AN model was designed for the holistic management of intangibles in organizations, allowing to establish cause-effect relationships in enterprise environments when intangible variables are present such as emotion and sentiment. It consists of 5 layers: Management, Experiences, Emotions, Attitudes and Business that connects the intangible and tangible variables of a company [35], [120].

As we can appreciate in figure A3, the GE²AN model is designed to be interpreted from two perspectives at the same time. On the one hand, the consumer perspective is an abstraction of the abovementioned consumer decision process, in which their external stimuli are processed with inner beliefs, causing a purchase intention which in turn causes the consumer's behavior.

On the other hand, the Business perspective shows that corporations make management decisions which become experiences available to their prospective customers, with the objective of causing emotions towards their business, products and brands that influence sales and other business figures.

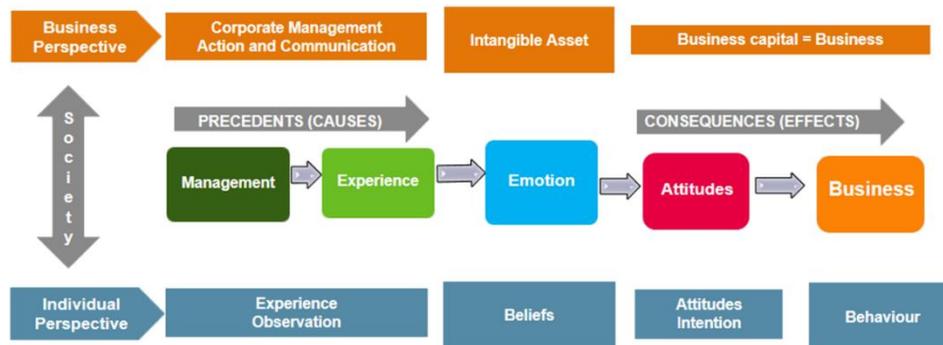


Figure A3: GE²AN intangibles management model [36].

The business' perspective of the GE²AN model is especially useful for companies because it provides a way to quantify intangible assets by establishing a relationship between them and the more easily quantifiable tangible assets. The first layer of this model is the company management, which encompasses all the information related to the decisions made by the company. Since management decisions are made within the corporation, this information is available to the company usually as internal information in the form of business reports and logs. The same applies to the last layer of the model, the business layer, which includes the business results of the company in the form of sales figures, income and other economic indicators.

On the other hand, experience, emotion and attitudes conform a sub-model of the GE²AN framework named E²A which describe the behavior of the purely intangible assets of the model. This sub-model presents some challenges when being measured as information on these is not readily available for the company and more often than not it has to be acquired from external sources such as consumer surveys and digital ecosystems. Furthermore, since the information acquired from these sources is heterogeneous in nature and mostly expressed in natural language, it must be pre-processed in order to obtain measurable data that can explain the relationship between management, consumers and business results within the role of intangibles management. This can be achieved by modelling the E²A sub-model as a multi-criteria decision making (MCDM) process.

A.2. Information sources on consumer behavior

One of the first steps for modeling consumer behavior is to obtain information from the consumer themselves. For this purpose, we can distinguish between two types of information according to the acquisition methods, structure and the techniques used to process them: solicited information and unsolicited information.

On the one hand, solicited information is obtained directly by polling users using surveys, focus groups and interviews in which the public answer directly to questions defined by researchers, being completely conscious of the questions and their purpose. On the other hand, unsolicited information is generated regularly and voluntarily by users without prompt. In this second category we find information from Social Media and digital ecosystems, metadata produced by regular web browsing and information on reservations and preorders among others. Due to the differences in the two types of information, nowadays most of the studies are incapable of working with both at the same time and are centered in one of these two, although each of them has its advantages and disadvantages.

The main advantage of solicited information is that, given its nature, an information retrieval campaign can be conducted on almost any topic, without depending on the online presence of said topic, which makes solicited information techniques a good tool to get data. The main characteristic of these techniques is that people who are being polled are aware of this information being retrieved, and they usually have at least a general idea of its intended uses when they are answering the questions. This can be an advantage since the information retrieval process can directly focus in the most important points, however, this awareness of the interviewee can lead to a strategic aspect of decisions, making them to give answers according to an objective or tending to the most socially acceptable options instead of exposing their true opinion, especially when dealing with sensitive topics [37].

Even though there are multiple solicited information retrieving techniques, the most used one are surveys. A survey consists of the development and deployment of a questionnaire with the intent to extract information from a particular group of people. A representative sample of the population is selected to answer the questionnaire and the information for the entire population is extrapolated from it. The creation of questionnaires must take into account several factors that can introduce bias such as the answers such as question and option order, the usage of open and closed questions, the sample size, the duration of the survey and how it will be delivered to the people who have to answer them. These aspects can have great effects in both the quality of the retrieved data and the deployment costs.

On the other hand, unsolicited information is obtained from online social actors who generate conversations and opinions without inciting them directly. The most usual way of obtaining this information is through the observation of organic conversations about different topics which are very usual in digital ecosystems. Working with data from digital ecosystems has its advantages such as the possibility of obtaining publicly available real-time data as it is generated on the Internet about almost any topic.

Most consumers who interact with products and services write opinions about them, detailing their experiences and emotions, which can be processed to obtain information about the intangible assets of businesses. Furthermore, this information is written spontaneously by the publics about the themes that are interesting to them, which allows us to avoid the strategic aspect of decisions [47].

The downside of unsolicited information is the variability in quality of the obtained data. There are methods that allow us to measure this quality, but their use depends on the study since when working with unsolicited information we are mostly searching for global tendencies. In these cases, the loss of precision assumable due to the sheer quantity of the data available [49].

A.3. Online Reputation Indices

Online reputation indices (ORIs) are proxy measures of the value of an intangible asset which aim to determine the value of an entity through opinion information obtained from the publics. Currently, they are one of the most widely used prescription source for consumers online for them to form an opinion on products and services. Online sites such as Amazon, eBay and TripAdvisor use scales such as stars, likes and valuations which allow prospective consumers to know, at a glance, the opinion of past customers, transforming them in social prescribers of said products and services [58], [59].

One of their most celebrated characteristics is their capability of conveying an opinion in just a number or a tag, especially when they aggregate multiple opinions, as they allow us to get a general idea of the degree of satisfaction of other customers and what to expect. However, the summary of this information causes an unavoidable loss of information, which can mislead the expectations of customers. In reality, there are multiple factors that affect the consumer when making a decision which is why, even though ORIs are key for understanding how they perceive products, more powerful tools are needed to assess the whole decision process.

A.4. Multi Criteria Decision Making

Multi Criteria Decision Making (MCDM) is a branch of operational research models which deal with decision making problems. A decision problem consists of identifying the most suitable option between a set of possible alternatives, by ordering the priorities of people who have knowledge about the goal according to their criteria. These criteria are often conflicting between decision makers, so the final solution tries to satisfy the majority of their preferences.

In our daily live we perform the process of weighing multiple criteria and comparing multiple alternatives instinctively. However, when there are complex decision problems with multiple decision-makers and competing interests, a more formal approach is desirable for reaching a compromise. MCDM methods such as analytic hierarchy process (APH), analytic network process (ANP), fuzzy set theory, ELECTRE and PROMETHEE were conceived with the goal of identifying the most desirable solutions for a decision problem in these situations [69].

These methods have been widely used in consumer research for determining the weights of evaluation criteria and the existing preferences on products and services available in the market in different situations. Different multi-criteria approaches have been successfully employed to evaluate new markets, choosing provider brands, optimizing workflow, recruit new personal and measuring consumer satisfaction among other things. In these works, the usage of MCDM techniques, has led to model complex business processes with some great results in understanding consumer behavior within their particular field [70]–[77]. Nonetheless, since there are multiple MCDM techniques, each of this scientific works had to choose the most suitable method and it had to be carefully adapted to their particular problem, along with experts extracting information about the criteria and alternatives being evaluated. The reason for there being many MCDM methods is Arrow's impossibility theorem which makes impossible to create a universal multi-criteria decision-making approach that can satisfy every decision problem at the same time.

A.4.1. Arrow's impossibility theorem

The main objective of any decision-making process is to select the most desirable option among a set of several possible alternatives. Furthermore, when there are multiple decision-makers, the decision-making process has the added task of ensuring the fairness of the collective decision, for this reason, Arrow [78] defines a set of five conditions or axioms that the resulting rank order must meet to truly reflect the opinions of every decision-maker:

- Pareto efficiency: If any option is preferred by everyone, this preference shall be preferred in the final ordering. This means that, given any two options A and B , if every decision maker states a preference of $A > B$ then the aggregated ranking must preserve said preference.
- Non-Dictatorship: The decision-making process is not a dictatorship, meaning that no individual decision-maker enjoys a position in which their preference must be preserved in the final ranking regardless of preferences expressed by others.
- Transitivity: The order of the final ranking is consistent, in which if $A > B$ and $B > C$, then it must be true that $A > C$.
- Unrestricted domain: Decision-makers have no restrictions in their individual order of alternatives. The aggregated resulting ranking should be unique and deterministic.
- Independence of irrelevant alternatives: Introducing or eliminating an option must not produce a rank reversal on the other options. That is, given a ranking in the form of $\{A, B\}$, expanding the options with a new option X , the resulting ranking must preserve the preference of $A > B$. Moreover, given a ranking in the form of $\{A, B, C\}$, making option B unavailable must preserve the preference of $A > C$.

Although all the above axioms must be satisfied by any MCDM in order to ensure fairness in the decision process, Arrow's impossibility theorem proves that it is impossible to satisfy all of them at the same time [79]. The consequence of this demonstration is that it is not possible to create a single decision-making method that can be used and is fair in each possible decision making problem, leading to one of the main criticism of MCDM methods: even with the same input, they may produce different and potentially conflicting rankings.

Some previous works have tried to tackle this consequence of Arrow's impossibility theorem, by using different aggregation methods and trying to combine results from multiple MCDM into a single ranking [80]. However, the creation of a super-ranking merging together results from other rankings results in a ranking itself, which is also affected by Arrow's impossibility theorem. The most extended solution to this problem is to relax some of Arrow's axioms in order to attain a good enough solution for any given particular problem. In our case, this problem being the determination of consumer preferences during the alternatives valuation phase of the purchasing decision process.

This presents additional challenges, since currently available methods were designed to be used in controlled environments in which decision-makers are reachable in order to clarify their statements, but this is not possible when using information obtained from digital ecosystems, especially when dealing with unsolicited information. For this reason, we need to

develop new MDCM methods explicitly designed to model the preferences of consumers, which also tackle the main factors affecting consumer decisions.

A.5. Factors affecting consumer decisions

When a consumer finally decides to purchase a good or service, many factors are involved. When modeling this behavior as an MCDM process, the different competing products available to the consumer are modeled as alternatives and the factors affecting the decision process are treated as the criteria of the multi-criteria decision-making model. However, determining and valuing criteria linked to the purchase decision process is a complex task.

Even if the stimuli that influence consumers can be observed from interactions in digital ecosystems, a large part of the stimuli are internal and imbued in the consumer's memory, being affected by their knowledge and experiences, which makes this measurement difficult. The various channels available for obtaining information present an additional challenge, since in order to understand what criteria consumers really take into account when choosing a product or service, it is necessary to study their opinions in all possible forms, regardless of whether this appears in the form of solicited or unsolicited information.

A.5.1. The combination of solicited and unsolicited information

The main advantage of solicited information is that subjects can be consulted directly about the areas of interest of the study so that, if the correct methods are used, we can obtain structured and much more accurate information than that obtained from unsolicited information. However, there are certain problems that we encounter when using solicited information. First, this type of information is generally more difficult to obtain, so the samples obtained tend to be much smaller than when working with unsolicited information. Secondly, the creation of questionnaires is a task that requires specialized personnel in the studied field as well as statistical and psychological knowledge if we want to avoid the many biases which can be introduced as a result of sampling errors, question design and other elements capable of indirectly influence respondents' answers such as the strategic aspect of decisions [104]–[106]. All these aspects, if not adequately treated, can reduce the validity of the data, so it is necessary to take measures to avoid them during the elaboration of the survey.

On the other hand, unsolicited information comes from conversations in natural language, generally from comments made on the Internet without the interlocutors having been asked specific questions, thus avoiding the strategic aspect of the decisions, greatly reducing certain biases and making it possible to obtain sufficiently large samples to mitigate sampling errors as much as possible or even to work with the entire population. However, in these cases, the researcher is relegated to a passive role, since he does not intervene and is limited to observing the behavior of consumers in their natural environment, so that the questions answered may not be those that the researcher really wants to know.

Ideally, we would use a hybrid approach in which solicited and unsolicited information can be combined, so that both types of data can reinforce each other and provide a much more global view of what the consumer thinks when making a purchase. Although the combination of both types of information is a great step towards the understanding of this process, it would not be sufficient to obtain a general and complete picture of the consumer since there are other problems such as synergies between consumer criteria and the perception of the alternatives' values which would also have to be tackled.

A.5.2. Synergies between consumer criteria

Once information is available, the next step is to use it to extract the characteristics or criteria on which consumers base their decision. There are many papers in the scientific literature that use MCDM techniques to extract information on the importance of criteria in different categories of products and services. These studies, which are usually designed specifically for a particular industry, start mainly from requested information and process it with the help of a panel of experts to identify the criteria used by consumers and the importance they assign internally to these criteria when choosing between one product or another [72], [76], [107]–[109].

However, the works on this topic have a tendency to consider each of the criteria as independent product characteristics, ignoring the synergies between them that are present in the consumer's mind when evaluating products. These interactions between criteria, when they are evaluated, are often the determining factor in the choice between several products and therefore, they must be considered when modeling consumer behavior. For example, a customer must choose between four products (Products A, B, C and D) that are evaluated on a scale [0,10], 0 being worst and 10 being the best, with respect to three criteria (design, quality and pricing) as shown in figure A4.

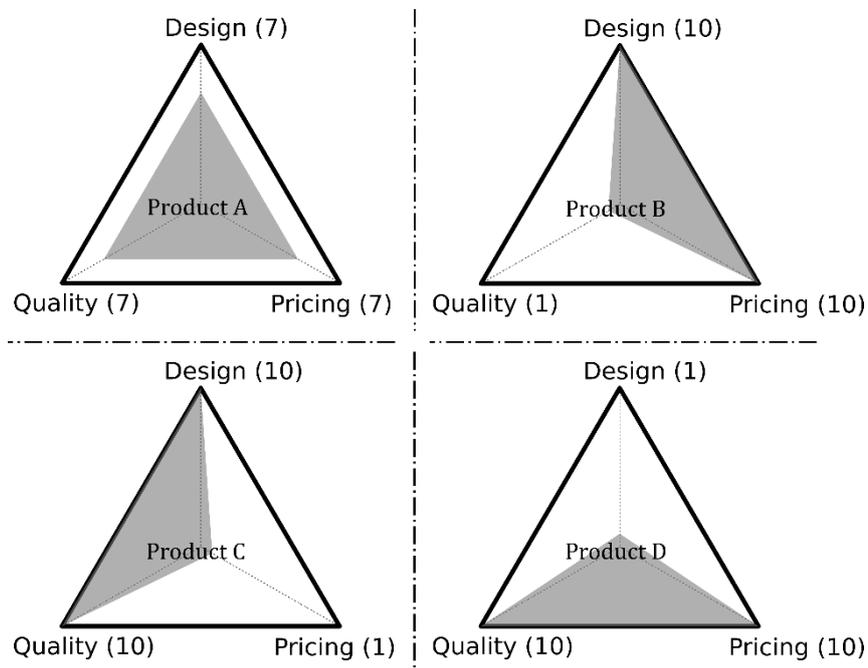


Figure A4: Valuation of four products with regards to design, quality and price

Obviously, when choosing between products like these, other factors such as the customer's previous experiences and emotions, as well as the motivation and intended use of the product, should be taken into account, but to simplify the problem let us assume that, for this particular customer, all criteria are of equal importance. In that case, one would think that since for each of the four products present their ratings add up to exactly 21 out of 30, every alternative should be equally attractive.

However, at a glance, product A is more attractive than the three others because of the synergies between the design, pricing and quality criteria chosen for this example. It is precisely this synergy between criteria which makes product A more desirable, as it provides an appropriate combination of "good, nice and cheap" than to discard one of them in favor of the other two.

When modeling this type of problem as a MCDM, when there is no relationship between the criteria, the importance between different sets of criteria is considered to be additive. This means that when a decision is made between any two criteria c_1 and c_2 together with the weights $w(c_1)$ and $w(c_2)$ that have been extracted to solve the decision problem, the weight of the combined criteria is given by:

$$w(c_1, c_2) = w(c_1) + w(c_2) \quad (A1)$$

However, this equation is only applicable when considering the criteria as independent entities. When consumers think about the alternatives, the criteria they use to evaluate them are interrelated to each other, so there is actually a value $d \in \mathbb{R}$ that affects the interaction between the different criteria, leading to the true weight of the combined criteria being:

$$w(c_1, c_2) = w(c_1) + w(c_2) + d \quad (A2)$$

Depending on the value of d we have three possibilities:

- If $d = 0$, criteria are considered independent and the relationship between them is considered additive.
- If $d > 0$, criteria c_1 and c_2 gain importance when they are present together. These criteria are more desirable when as a group than separated, reinforcing themselves in these situations, such as the aforementioned example of quality and attractive pricing.
- If $d < 0$, criteria c_1 and c_2 overlap, therefore the importance of them being together is less than the sum of the weights of both criteria by themselves. As an example, when a brand is well known and appreciated by the public, there is usually an implicit level of quality, which makes situations in which these two characteristics are not presented together more important than situations in which these two criteria are.

Even though this phenomenon has been studied in the scientific literature, the only solution available so far is to extract information using groups of expert decision-makers [110]–[114]. This approach, although valid for use in decision making problems in small groups with solicited information, is insufficient for modeling the behavior of third parties based on unsolicited opinions resulting from a Data Mining process. The methodology proposed in this work enables us to directly extract criteria synergies from large sets of communications, written by real consumers who give their opinion about products and services, and allow us to approximate in a more realistic way the value that they give to the products and services they purchase.

A.5.3. The value of comparisons

One aspect in common with every model attempting to explain consumer behavior within the purchasing decision process is the valuation of possible alternatives using a scale. This is a logical step since the valuation of products by the consumer is one of the five stages of commercial transactions. However, this step is relatively complex since the value of a product or service is a subjective characteristic that each individual assigns differently. In addition, psychological aspects unique to each individual come into play during this value assignment, which can lead to decisions that, a priori, may appear irrational [115].

Currently, the most widely used tool to approximate the value of a product based on information from digital ecosystems are the ORIs explained above. Although ORIs do not indicate the value of the product, these indicators are used as a proxy measure of the product, giving us a general idea of the value of a product. A product with a very low valuation has a high probability of being bad, and another one with a very high valuation has a high probability of being a good solution to the problem it is intended to solve.

Traditional ORIs work well for obtaining information about products in isolation. However, products hardly exist as isolated entities, but coexist in a market where multiple competing products compete with other alternative solutions, so the perceived value of a product cannot be measured as an isolated value either without accounting the comparisons that consumers constantly make between the different alternatives.

This is the reason why the currently used ORIs, such as stars, ratings and number of likes, are not as useful to measure value as desired and some digital platforms go so far as to stop displaying them. There are really two major problems that undermine the usefulness of traditional ORIs as rating systems: a) The difference in the internal scale between different users and b) The dissociation between ORIs and the concept of comparison.

On the one hand, information from multiple decision makers is being aggregated into a single value, when internally each of them uses a different valuation scale. In other words, when issuing a valuation, the “★★★★☆” given by two different users rarely have the same meaning for both of them. On the other hand, there is a strong dissociation between the way ORIs are built and the concept of comparison inherent to the valuation process performed by consumers on the alternatives.

The second problem, the dissociation between ORIs and the concept of comparison, is independent of the type of information used and is more related to the inner workings of ORIs and their use as a proxy measure of a product's value in decision making. When someone looks

at the rating given to a product on a website, what they really want to get is information about the inherent value of that product to other people. This approach is incorrect because the value of a good is not an isolated characteristic of the object, but the result of a series of comparisons of that same good against all similar products.

The process of calculating an ORI designed to approximate the value that the public places on a product should be analogous to human behavior. When valuing any entity, what a person actually does is to make a series of comparisons with other similar entities that he knows and can serve as a basis for building an idea of how things should be and to be able to establish this valuation. In other words, the value of a good depends directly on the comparison with other goods of the same class [116].

Therefore, if we really want to model consumer behavior based on information from digital ecosystems, it is essential to establish an indicator based on this comparison of sets of comparable alternatives that can be used as a measure of the value of a product. To do this, it is necessary to build an ORI that works on the basis of comparisons of the product in its context. This context is the set of comparable alternatives in which the product coexists.

A.6. Conclusions

In this work we have analyzed the problems of modeling consumer behavior based on information from digital ecosystems, detecting existing weaknesses in the methodology currently used and reinforcing them by providing complementary methods that can solve these problems in a more efficient and accurate way. After studying the current state of the art, three major problems have been detected that make it difficult to obtain knowledge about consumer behavior, all of them appreciable during the evaluation phase of alternatives that the consumer carries out when making a purchase.

On the one hand, the information that was being used up to now was incomplete, due to the differences in the characteristics of the data provided by different sources. There was no way of adequately combining the solicited information with unsolicited information. In chapter 2 of this work, a methodology has been established to combine both types of information obtained through the analysis of digital ecosystems, with data obtained from traditional demographic techniques. The combined use of this complementary information has demonstrated an appreciable improvement in the quality of the knowledge that can be obtained about the decisions made by consumers.

On the other hand, when generating these rankings, the interactions between the criteria that consumers internally use when evaluating products presented another unresolved challenge that, until now, had made it difficult to understand why consumers choose certain products over others. After an exhaustive review of the scientific literature, it can be seen that no methodology was able to extract these naturally occurring synergies between criteria, quantify them and use them to improve purchase intent prediction. The new technique for calculating criteria synergies provided in chapter 3 of this work allows, in an automatable way, and based on the implicit relationships in valuations expressed by the public, to quantify the effect of criteria synergies and use them to build rankings of purchasable alternatives.

With the inclusion of data on synergies, the new methodology provided has shown not only that the presence of criteria synergies in purchase decisions affects the final ranking of alternatives, but also that considering the effect of these synergies brings the results of predictive purchase rankings even closer to the real sales values of products when they enjoy a certain level of presence in digital ecosystems.

Finally, it has been detected that the most widespread use of traditional ORIs, i.e., as a proxy measure of the appreciated value of a product by audiences, entailed a series of problems derived both from the different internal scales used by people to value, and the dissociation of the concept of "value" of a product with the valuation of comparable alternatives. This means that measures such as stars or likes on a product are hardly ever useful for comparing between products. Although thanks to the methodology provided in chapter 3 we can extract information on the weight of criteria when making a decision, the perceived value of a product is obfuscated by this interpretation of traditional ORIs as a proxy measure.

In order to obtain more realistic information about the perceived value of products within their market, a new ORI has been developed, presented in chapter 4, which is built from scratch based on the functioning of the psychological processes that consumers use to assign value to products. That is, based on comparisons between groups of comparable alternatives. The CMGI allows to extract this information directly from opinions expressed in natural language, as the public would do, and following a fully automatable and deterministic methodology, to produce a ranking that reflects the relative appreciation that the public has for each of the available alternatives.

Although these three methodologies can be used in a stand-alone fashion, their true potential appears when they are used in combination to model the evaluation of alternatives phase, which is present in every purchase decision process. By using these three techniques together, it is possible to extract the criteria that consumers use, including the synergies implicit in them, and to know the value that the public places on the available products, all this using the

maximum amount of information possible thanks to a combination of solicited and unsolicited information.

One of the biggest problems caused by the rapid growth of digital ecosystems is the huge amount of weakly structured information available, presenting itself as an amalgam of data that must be processed by companies and institutions if they really want to know what audiences demand from them. Knowing the criteria and alternatives on which these audiences base their consumption decisions is an important step to bring some order to all this chaos and will be a key piece for an increasingly communicated future.

Bibliografía

- [1] C. Nwabueze and E. Okonkwo, “Rethinking the Bullet Theory in the Digital Age,” *International Journal of Media, Journalism and Mass Communications*, vol. 4, no. 2, pp. 1–10, 2018, doi: 10.20431/2454-9479.0402001.
- [2] J. P. Martino, *Technological forecasting for decision making*. 1993.
- [3] A. M. Casado and J. I. Peláez, “Intangible management monitors and tools: Reviews,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1509–1529, Mar. 2014, doi: 10.1016/j.eswa.2013.08.048.
- [4] M. McGuirk, “Performing social media analytics with Brandwatch for Classrooms: a platform review,” *Journal of Marketing Analytics 2021*, pp. 1–16, Aug. 2021, doi: 10.1057/S41270-021-00128-5.
- [5] J. Clarkson and L. Abbiss, “Behavioural transitions of bottlenose dolphins Montenegro, South Adriatic,” *2018 IEEE International Workshop on Metrology for the Sea*, pp. 129–134, 2018.
- [6] L. E. Long, C. Leung, J. S. Hong, C. Wright, and C. J. Young, “Patterns of internet and social media use in colorectal surgery,” *BMC Surgery*, vol. 19, no. 1, p. 52, Dec. 2019, doi: 10.1186/s12893-019-0518-4.
- [7] Conviva, “Conviva: Viewer Experience Report,” Foster City, 2014. [Online]. Available: <https://www.conviva.com/newsroom/conviva-releases-2014-viewer-experience-report/>
- [8] P. Sahatiya, “Big Data Analytics on Social Media Data: A Literature Review,” *International Research Journal of Engineering and Technology*, vol. 5, no. 2, pp. 189–192, 2018.
- [9] Mcdaniel & Gates, “Marketing research. International student version,” *John Wiley & Sons*, 2018, [Online]. Available: <http://doc1.lbfli.li/aca/FLMF037627.pdf>
- [10] L.-T. (Jane) Hsu and S. (Shawn) Jang, “Advertising expenditure, intangible value and risk: A study of restaurant companies,” *International Journal of Hospitality Management*, vol. 27, no. 2, pp. 259–267, Jun. 2008, doi: 10.1016/j.ijhm.2007.07.012.

- [11] L. J. Parsons and F. M. Bass, “Optimal Advertising-Expenditure Implications of a Simultaneous-Equation Regression Analysis,” vol. 19, no. 3, pp. 822–831, 1971.
- [12] H. Boley and E. Chang, “Digital Ecosystems: Principles and Semantics,” in *2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference*, Feb. 2007, pp. 398–403. doi: 10.1109/DEST.2007.372005.
- [13] Z. Liao, “Environmental policy instruments, environmental innovation and the reputation of enterprises,” *Journal of Cleaner Production*, vol. 171, pp. 1111–1117, Jan. 2018, doi: 10.1016/j.jclepro.2017.10.126.
- [14] A. Akhigbe, J. E. McNulty, and B. A. Stevenson, “How does transparency affect bank financial performance?,” *International Review of Financial Analysis*, vol. 29, pp. 24–30, Sep. 2013, doi: 10.1016/j.irfa.2013.01.007.
- [15] P. J. Kitchen and I. Burgmann, “Integrated Marketing Communication,” in *Wiley International Encyclopedia of Marketing*, vol. 18, no. 5, Chichester, UK: John Wiley & Sons, Ltd, 2010, pp. 613–626. doi: 10.1002/9781444316568.wiem04001.
- [16] U. W. Chohan, “Counter-Hegemonic Finance: The Gamestop Short Squeeze,” *SSRN Electronic Journal*, no. January, 2021, doi: 10.2139/ssrn.3775127.
- [17] C. Long, B. M. Lucey, and L. Yarovaya, “‘I Just Like the Stock’ versus ‘Fear and Loathing on Main Street’: The Role of Reddit Sentiment in the GameStop Short Squeeze,” *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3822315.
- [18] E. P. Barinova, E. N. Sheremetyeva, and A. S. Zotova, “Digital Talents: Realities and Prospects,” in *Lecture Notes in Networks and Systems*, vol. 84, 2020, pp. 327–334. doi: 10.1007/978-3-030-27015-5_39.
- [19] N. Shen, K. Au, and W. Li, “Strategic alignment of intangible assets: The role of corporate social responsibility,” *Asia Pacific Journal of Management*, Aug. 2019, doi: 10.1007/s10490-019-09681-1.
- [20] S. C. Lim, A. J. Macias, and T. Moeller, “Intangible assets and capital structure,” *Journal of Banking & Finance*, vol. 118, p. 105873, Sep. 2020, doi: 10.1016/j.jbankfin.2020.105873.
- [21] Y. Wang and G. Ruhe, “The Cognitive Process of Decision Making,” *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 1, no. 2, pp. 73–85, Apr. 2007, doi: 10.4018/jcini.2007040105.

- [22] G. R. Foxall, "Understanding consumer choice," *Understanding Consumer Choice*, pp. 1–262, 2005, doi: 10.1057/9780230510029.
- [23] P. Kotler, *Marketing Management: The Millennium Edition*. UpperSaddle River, NJ: Prentice-Hall, 2000.
- [24] R. Raffaelli, "Technology Reemergence: Creating New Value for Old Technologies in Swiss Mechanical Watchmaking, 1970–2008," *Administrative Science Quarterly*, vol. 64, no. 3, pp. 576–618, 2019, doi: 10.1177/0001839218778505.
- [25] E. Vidler and J. Clarke, "Creating Citizen-Consumers: New Labour and the Remaking of Public Services," *Public Policy and Administration*, vol. 20, no. 2, pp. 19–37, 2005, doi: 10.1177/095207670502000202.
- [26] M. Lavin, "Creating Consumers in the 1930s: Irna Phillips and the Radio Soap Opera," *Journal of Consumer Research*, vol. 22, no. 1, p. 75, 1995, doi: 10.1086/209436.
- [27] D. Voramontri and L. Klieb, "Impact of social media on consumer behaviour," 2019.
- [28] B. Verplanken and W. Wood, "Interventions to Break and Create Consumer Habits," *Journal of Public Policy & Marketing*, vol. 25, no. 1, pp. 90–103, Apr. 2006, doi: 10.1509/jppm.25.1.90.
- [29] M. Racine, C. Wilson, and M. Wynes, "The Value of Apology: How do Corporate Apologies Moderate the Stock Market Reaction to Non-Financial Corporate Crises?," *Journal of Business Ethics*, vol. 163, no. 3, pp. 485–505, 2020, doi: 10.1007/s10551-018-4037-5.
- [30] J. Wei, Z. Ouyang, and H. A. Chen, "Well Known or Well Liked? The Effects of Corporate Reputation on Firm Value at the Onset of a Corporate Crisis," *Strategic Management Journal*, vol. 38, no. 10, pp. 2103–2120, Oct. 2017, doi: 10.1002/smj.2639.
- [31] R. R. Dolphin, "Corporate reputation – a value creating strategy," *Corporate Governance: The international journal of business in society*, vol. 4, no. 3, pp. 77–92, 2004, doi: 10.1108/14720700410547521.
- [32] S.-P. Jeng, "Effects of corporate reputations, relationships and competing suppliers' marketing programmes on customers' cross-buying intentions," *The Service Industries Journal*, vol. 28, no. 1, pp. 15–26, Jan. 2008, doi: 10.1080/02642060701725370.

- [33] K. J. Zink, “Stakeholder orientation and corporate social responsibility as a precondition for sustainability,” *Total Quality Management & Business Excellence*, vol. 16, no. 8–9, pp. 1041–1052, Oct. 2005, doi: 10.1080/14783360500163243.
- [34] Fundación Corporate Excellence, “Informes de Intangibles INTED 2018,” Madrid, 2018. [Online]. Available: <https://www.corporateexcellence.org/recurso/iii-informe-de-activos-intangibles-en-ecosistemas/e7dfa484-555b-4a4a-bf84-9cafc410ac10>
- [35] J. I. Peláez, G. Vaccaro, and F. E. Cabrera, “The Current Status of Social Risks on Educational Systems. an Analysis Through Social Media,” *ICERI2018 Proceedings*, vol. 1, no. November, pp. 2374–2381, 2018, doi: 10.21125/iceri.2018.1522.
- [36] J. I. Pelaez, A. Casado, E. R. Yanez, F. E. Cabrera, and A. H. H. Yanez, “A model for social assessment of governmental areas through intangible assets,” in *2017 4th International Conference on eDemocracy and eGovernment, ICEDEG 2017*, Apr. 2017, pp. 200–204. doi: 10.1109/ICEDEG.2017.7962534.
- [37] R. Tourangeau and T. Yan, “Sensitive Questions in Surveys,” *Psychological Bulletin*, vol. 133, no. 5, pp. 859–883, 2007, doi: 10.1037/0033-2909.133.5.859.
- [38] P. Hartley, P. W. Routon, and L. Torres, “The Skills Marketing Majors Believe They Acquire: Evidence From a National Survey,” *Journal of Marketing Education*, vol. 41, no. 3, pp. 202–214, Dec. 2019, doi: 10.1177/0273475318757282.
- [39] Z. T. Plakias, I. Demko, and A. L. Katchova, “Direct marketing channel choices among US farmers: evidence from the Local Food Marketing Practices Survey,” *Renewable Agriculture and Food Systems*, vol. 35, no. 5, pp. 475–489, Oct. 2020, doi: 10.1017/S1742170519000085.
- [40] F. R. Santos, “Satisfacción, bienestar y calidad de vida en el trabajo,” *Reis*, no. 92, p. 11, 2000, doi: 10.2307/40184292.
- [41] P. Fanlo *et al.*, “Impacto de la infección por el nuevo coronavirus en los pacientes con uveítis asociada a una enfermedad autoinmune: resultado de la encuesta COVID19-GEAS pacientes,” *Archivos de la Sociedad Española de Oftalmología*, Jan. 2021, doi: 10.1016/j.oftal.2020.12.011.
- [42] A. Alaminos Chica and J. L. Castejón Costa, *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión*. Alicante, Spain: Editorial Marfil, S.A., 2006. doi: 10.1097/01.ccm.0000457813.83878.83.

- [43] K. L. Manfreda, M. Bosnjak, J. Berzelak, I. Haas, and V. Vehovar, “Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates,” *International Journal of Market Research*, vol. 50, no. 1, pp. 79–104, Jan. 2008, doi: 10.1177/147078530805000107.
- [44] A. Mavletova, “Data Quality in PC and Mobile Web Surveys,” *Social Science Computer Review*, vol. 31, no. 6, pp. 725–743, Dec. 2013, doi: 10.1177/0894439313485201.
- [45] W. B. Meigs, “The Expanding Field of Internal Auditing,” *The Accounting Review*, vol. 26, no. 4, pp. 518–523, 1951.
- [46] B. Batrinca and P. C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms,” *AI and Society*, vol. 30, no. 1, pp. 89–116, 2014, doi: 10.1007/s00146-014-0549-4.
- [47] S. Balbi, M. Misuraca, and G. Scepi, “Combining different evaluation systems on social media for measuring user satisfaction,” *Information Processing and Management*, vol. 54, no. 4, pp. 674–685, 2018, doi: 10.1016/j.ipm.2018.04.009.
- [48] J. I. Peláez, E. A. Martínez, and L. G. Vargas, “Decision making in social media with consistent data,” *Knowledge-Based Systems*, vol. 172, pp. 33–41, May 2019, doi: 10.1016/J.KNOSYS.2019.02.009.
- [49] K. Mayer-Schönberger, Viktor & Cukier, *Big Data. La revolución de los datos masivos*. 2013.
- [50] J. R. Evink, “Information Ecology: Mastering the Information and Knowledge Environment Information Ecology: Mastering the Information and Knowledge Environment By Davenport Thomas H., New York: Oxford University Press, 1997. 229 pages, hard cover, \$29.95,” *Academy of Management Perspectives*, vol. 11, no. 4, pp. 98–100, Nov. 1997, doi: 10.5465/ame.1997.9712024845.
- [51] C. L. Philip Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Information Sciences*, vol. 275, pp. 314–347, 2014, doi: 10.1016/j.ins.2014.01.015.
- [52] R. M. Losee, M. Hall, U. N. C. Hill, and C. Hill, “Browsing Mixed Structured and Unstructured Data Information Processing & Management,” *Information Processing & Management*, vol. 42, no. March, pp. 440–452, 2006.

- [53] S. Vyas and P. Vaishnav, “A comparative study of various ETL process and their testing techniques in data warehouse,” *Journal of Statistics and Management Systems*, vol. 20, no. 4, pp. 753–763, 2017, doi: 10.1080/09720510.2017.1395194.
- [54] S. Gorhe, “ETL in Near-Real Time Environment : Challenges and Opportunities,” no. April, 2020.
- [55] F. Rubio, P. Vega, and R. P. R. Ch, “NoSQL vs . SQL in Mass Data Management : An Empirical Study NoSQL contra SQL en la Administración de datos masivos : Un estudio empírico,” vol. 2020, pp. 40–49, 2020, doi: 10.18502/keg.v5i1.5917.
- [56] C. Murdough, “Social Media Measurement: It’s Not Impossible,” *Journal of Interactive Advertising*, vol. 2019, 2019, doi: 10.1080/15252019.2009.10722165.
- [57] S. E. Seker, “Information and Communication Technology Reputation for XU030 Quote Companies,” *International Journal of Innovation, Management and Technology*, vol. 5, no. 3, pp. 221–225, 2014, doi: 10.7763/IJIMT.2014.V5.517.
- [58] M. Hu and B. Liu, “Mining and summarizing customer reviews,” *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, p. 168, 2004, doi: 10.1145/1014052.1014073.
- [59] M. Luca, “Reviews, Reputation, and Revenue: The Case of Yelp.Com,” *SSRN Electronic Journal*, 2011, doi: 10.2139/ssrn.1928601.
- [60] Y. Han and K. K. Kim, “Sentiment analysis on social media using morphological sentence pattern model,” *Proceedings - 2017 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications, SERA 2017*, pp. 79–84, Jun. 2017, doi: 10.1109/SERA.2017.7965710.
- [61] T. Santos, F. Lemmerich, M. Strohmaier, and D. Helic, “What’s in a review: Discrepancies between expert and amateur reviews of video games on Metacritic,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, 2019, doi: 10.1145/3359242.
- [62] S. Suryawanshi and M. Narnaware, “Design and Analysis of Collaborative Filtering Based Recommendation System,” *International Journal of Engineering Applied Sciences and Technology*, vol. 5, no. 4, pp. 223–226, 2020, doi: 10.33564/ijeast.2020.v05i04.031.

- [63] B. Cankir, M. L. Arslan, and S. E. Seker, "Web Reputation Index for XU030 Quote Companies," *Journal of Industrial and Intelligent Information*, vol. 3, no. 2, pp. 2013–2016, 2015, doi: 10.12720/jiii.3.2.110-113.
- [64] M. Arslan, "Web Based Reputation Index of Turkish Universities," *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 4, no. 3, 2014, doi: 10.7763/ijeeee.2014.v4.330.
- [65] J. F. Perles-Ribes, A. B. Ramón-Rodríguez, L. Moreno-Izquierdo, and M. J. Such-Devesa, "Online reputation and destination competitiveness: The case of Spain," *Tourism Analysis*, vol. 24, no. 2, pp. 161–176, 2019, doi: 10.3727/108354219X15525055915518.
- [66] C. K. Anderson and B. Lawrence, "The Influence of Online Reputation and Product Heterogeneity on Service Firm Financial Performance," *Service Science*, vol. 6, no. 4, pp. 217–228, 2014, doi: 10.1287/serv.2014.0080.
- [67] X. Liu, X. Guo, H. Wu, and T. Wu, "The Impact of Individual and Organizational Reputation on Physicians' Appointments Online," *International Journal of Electronic Commerce*, vol. 20, no. 4, pp. 551–577, 2016, doi: 10.1080/10864415.2016.1171977.
- [68] B. Wansink and J. Sobal, "Mindless Eating," *Environment and Behavior*, vol. 39, no. 1, pp. 106–123, Jan. 2007, doi: 10.1177/0013916506295573.
- [69] M. Velasquez and P. T. Hester, "An Analysis of Multi-Criteria Decision Making Methods," *International Journal of Operations Research*, vol. 10, no. 2, pp. 56–66, 2013, doi: 10.1007/978-3-319-12586-2.
- [70] S. Aksoy and M. Yetkin Ozbuk, "Multiple criteria decision making in hotel location: Does it relate to postpurchase consumer evaluations?," *Tourism Management Perspectives*, vol. 22, pp. 73–81, 2017, doi: 10.1016/j.tmp.2017.02.001.
- [71] S. D. Pohekar and M. Ramachandran, "Application of multi-criteria decision making to sustainable energy planning - A review," *Renewable and Sustainable Energy Reviews*, vol. 8, no. 4, pp. 365–381, 2004, doi: 10.1016/j.rser.2003.12.007.
- [72] G. Işıklar and G. Büyüközkan, "Using a multi-criteria decision making approach to evaluate mobile phone alternatives," *Computer Standards & Interfaces*, vol. 29, no. 2, pp. 265–274, Feb. 2007, doi: 10.1016/j.csi.2006.05.002.
- [73] N. F. Matsatsinis and A. P. Samaras, "Brand choice model selection based on consumers' multicriteria preferences and experts' knowledge," *Computers and*

Operations Research, vol. 27, no. 7–8, pp. 689–707, 2000, doi: 10.1016/S0305-0548(99)00114-8.

- [74] N. Sousa, A. Almeida, and J. Coutinho-Rodrigues, “A multicriteria methodology for estimating consumer acceptance of alternative powertrain technologies,” *Transport Policy*, vol. 85, no. September 2019, pp. 18–32, 2020, doi: 10.1016/j.tranpol.2019.10.003.
- [75] E. Mulliner, K. Smallbone, and V. Maliene, “An assessment of sustainable housing affordability using a multiple criteria decision making method,” *Omega (United Kingdom)*, vol. 41, no. 2, pp. 270–279, 2013, doi: 10.1016/j.omega.2012.05.002.
- [76] H. Zaim, M. Ramdani, and A. Haddi, “Multi-criteria analysis approach based on consumer satisfaction to rank B2C E-commerce websites,” *SITA 2016 - 11th International Conference on Intelligent Systems: Theories and Applications*, 2016, doi: 10.1109/SITA.2016.7772260.
- [77] E. K. Zavadskas, J. Antucheviciene, and P. Chatterjee, “Multiple-criteria decision-making (MCDM) techniques for business processes information management,” *Information (Switzerland)*, vol. 10, no. 1, pp. 1–7, 2018, doi: 10.3390/info10010004.
- [78] L. A. Goodman and K. J. Arrow, *Social Choice and Individual Values.*, vol. 18, no. 1. Yale university press, 1953. doi: 10.2307/2087870.
- [79] C. Press, “A Difficulty in the Concept of Social Welfare Author (s): Kenneth J . Arrow Source : Journal of Political Economy , Vol . 58 , No . 4 (Aug . , 1950) , pp . 328-346 Published by : The University of Chicago Press Stable URL : <http://www.jstor.org/stable/1>,” vol. 58, no. 4, pp. 328–346, 2018.
- [80] M. Mohammadi and J. Rezaei, “Ensemble ranking: Aggregation of rankings produced by different multi-criteria decision-making methods,” *Omega (United Kingdom)*, vol. 96, p. 102254, 2020, doi: 10.1016/j.omega.2020.102254.
- [81] B. Ceballos, M. T. Lamata, and D. A. Pelta, “A comparative analysis of multi-criteria decision-making methods,” *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 315–322, 2016, doi: 10.1007/s13748-016-0093-1.
- [82] T. L. Saaty, “Decision making with the analytic hierarchy process,” vol. 1, no. 1, 2008, doi: 10.1504/IJSSci.2008.01759.

- [83] T. L. Saaty, “Decision making with the analytic hierarchy process,” *International Journal of Services Sciences*, vol. 1, no. 1, p. 83, 2008, doi: 10.1504/IJSSCI.2008.017590.
- [84] E. Triantaphyllou, “Two new cases of rank reversals when the AHP and some of its additive variants are used that do not occur with the multiplicative AHP,” *Journal of Multi-Criteria Decision Analysis*, vol. 10, no. 1, pp. 11–25, 2001, doi: 10.1002/mcda.284.
- [85] A. Calabrese, R. Costa, N. Levialedi, and T. Menichini, “Integrating sustainability into strategic decision-making: A fuzzy AHP method for the selection of relevant sustainability issues,” *Technological Forecasting and Social Change*, vol. 139, no. March 2018, pp. 155–168, 2019, doi: 10.1016/j.techfore.2018.11.005.
- [86] J. Jablonsky, “Measuring the efficiency of production units by AHP models,” *Mathematical and Computer Modelling*, vol. 46, no. 7–8, pp. 1091–1098, 2007, doi: 10.1016/j.mcm.2007.03.007.
- [87] P. T. Harker, “Incomplete pairwise comparisons in the analytic hierarchy process,” *Mathematical Modelling*, vol. 9, no. 11, pp. 837–848, 1987, doi: 10.1016/0270-0255(87)90503-3.
- [88] M. Karanik, L. Wanderer, J. A. Gomez-Ruiz, and J. I. Pelaez, “Reconstruction methods for AHP pairwise matrices: How reliable are they?,” *Applied Mathematics and Computation*, vol. 279, pp. 103–124, Apr. 2016, doi: 10.1016/j.amc.2016.01.008.
- [89] J. Aguarón and J. M. Moreno-Jiménez, “The geometric consistency index: Approximated thresholds,” *European Journal of Operational Research*, vol. 147, no. 1, pp. 137–145, 2003, doi: 10.1016/S0377-2217(02)00255-2.
- [90] J. I. Pelaez, E. A. Martinez, and L. G. Vargas, “Consistency in Positive Reciprocal Matrices: An Improvement in Measurement Methods,” *IEEE Access*, vol. 6, pp. 25600–25609, 2018, doi: 10.1109/ACCESS.2018.2829024.
- [91] J. I. Peláez and M. T. Lamata, “A new measure of consistency for positive reciprocal matrices,” *Computers and Mathematics with Applications*, vol. 46, no. 12, pp. 1839–1845, 2003, doi: 10.1016/S0898-1221(03)90240-9.
- [92] J. A. Gomez-Ruiz, M. Karanik, and J. I. Peláez, “Improving the Consistency of AHP Matrices Using a Multi-layer Perceptron-Based Model,” in *Bio-Inspired Systems: Computational and Ambient Intelligence*, 2009, pp. 41–48.

- [93] Z. Chen and D. Ben-Arieh, "On the fusion of multi-granularity linguistic label sets in group decision making," *Computers and Industrial Engineering*, vol. 51, no. 3, pp. 526–541, 2006, doi: 10.1016/j.cie.2006.08.012.
- [94] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-III," *Information Sciences*, vol. 9, no. 1, pp. 43–80, 1975, doi: 10.1016/0020-0255(75)90017-1.
- [95] Y. Liu, C. M. Eckert, and C. Earl, "A review of fuzzy AHP methods for decision-making with subjective judgements," *Expert Systems with Applications*, vol. 161, p. 113738, 2020, doi: 10.1016/j.eswa.2020.113738.
- [96] Marimin, M. Umano, I. Hatono, and H. Tamura, "Linguistic labels for expressing fuzzy preference relations in fuzzy group decision making," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 2, pp. 205–218, 1998, doi: 10.1109/3477.662760.
- [97] V. Cutello and J. Montero, "Hierarchies of intensity preference aggregations," *International Journal of Approximate Reasoning*, vol. 10, no. 2, pp. 123–133, 1994, doi: 10.1016/0888-613X(94)90012-4.
- [98] R. R. Yager, "Families of OWA Operators," *Fuzzy Sets Syst.*, vol. 59, no. 2, pp. 125–148, Oct. 1993, doi: 10.1016/0165-0114(93)90194-M.
- [99] J. M. Doña Fernández, "Modelado de los procesos de toma de decisión en entornos sociales mediante operadores de agregación owa," 2008.
- [100] R. R. Yager, "On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988, doi: 10.1109/21.87068.
- [101] J. Peláez and J. M. Doña, "Majority additive-ordered weighting averaging: A new neat ordered weighting averaging operator based on the majority process," *International Journal of Intelligent Systems*, vol. 18, no. 4, pp. 469–481, 2003, doi: 10.1002/int.10096.
- [102] M. Karanik, J. I. Peláez, and R. Bernal, "Selective majority additive ordered weighting averaging operator," *European Journal of Operational Research*, vol. 250, no. 3, pp. 816–826, 2016, doi: 10.1016/j.ejor.2015.10.011.

- [103] R. R. Yager, "Decision Making Under Dempster-Shafer Uncertainties," *International Journal of General Systems*, vol. 20, no. 3, pp. 233–245, 1992, doi: 10.1080/03081079208945033.
- [104] K. J. Arrow, *Social Choice and Individual Values*. New York, 1963.
- [105] A. Gibbard, "Manipulation of Voting Schemes," *Econometrica*, vol. 41, no. 4, pp. 587–601, 1973. doi: 10.2307/1914083.
- [106] H. Moulin, "On strategy-proofness and single peakedness," *Public Choice*, vol. 35, no. 4, pp. 437–455, 1980, doi: 10.1007/BF00128122.
- [107] G. F. Can and E. Kılıç Delice, "A task-based fuzzy integrated MCDM approach for shopping mall selection considering universal design criteria," *Soft Computing*, vol. 22, no. 22, pp. 7377–7397, Nov. 2018, doi: 10.1007/s00500-018-3074-4.
- [108] Y. Liu, Y. Yang, Y. Liu, and G. H. Tzeng, "Improving sustainable mobile health care promotion: A novel hybrid MCDM Method," *Sustainability (Switzerland)*, vol. 11, no. 3, pp. 1–29, 2019, doi: 10.3390/su11030752.
- [109] W. Poon and K. Lock-Teng Low, "Are travellers satisfied with Malaysian hotels?," *International Journal of Contemporary Hospitality Management*, vol. 17, no. 3, pp. 217–227, May 2005, doi: 10.1108/09596110510591909.
- [110] J. L. Yang, H. N. Chiu, G.-H. Tzeng, and R. H. Yeh, "Vendor selection by integrated fuzzy MCDM techniques with independent and interdependent relationships," *Information Sciences*, vol. 178, no. 21, pp. 4166–4183, Nov. 2008, doi: 10.1016/j.ins.2008.06.003.
- [111] S. Angilella, S. Greco, and B. Matarazzo, "Non-additive robust ordinal regression: A multiple criteria decision model based on the Choquet integral," *European Journal of Operational Research*, vol. 201, no. 1, pp. 277–288, Feb. 2010, doi: 10.1016/j.ejor.2009.02.023.
- [112] H. CHIOU, G. TZENG, and D. CHENG, "Evaluating sustainable fishing development strategies using fuzzy MCDM approach," *Omega*, vol. 33, no. 3, pp. 223–234, Jun. 2005, doi: 10.1016/j.omega.2004.04.011.
- [113] L. B. Reinhardt and D. Pisinger, "Multi-objective and multi-constrained non-additive shortest path problems," *Computers & Operations Research*, vol. 38, no. 3, pp. 605–616, Mar. 2011, doi: 10.1016/j.cor.2010.08.003.

- [114] R. Bernal, M. Karanik, and J. I. Peláez, “Fuzzy measure identification for criteria coalitions using linguistic information,” *Soft Computing*, vol. 20, no. 4, pp. 1315–1327, Apr. 2016, doi: 10.1007/s00500-015-1589-5.
- [115] F. Emuze and T. Saurin, *Value and Waste in Lean Construction*. London: Routledge, 2015. doi: 10.4324/9781315696713.
- [116] B. Andersson and P. Johannesson, “Ascribing exchange value,” *CEUR Workshop Proceedings*, vol. 2383, 2019.
- [117] F. E. Cabrera, M. Amaya, G. F. Vaccaro Witt, and J. I. Peláez, “Pairwise Voting to Rank Touristic Destinations Based on Preference Valuation,” *Sustainability*, vol. 11, no. 21, p. 5873, Oct. 2019, doi: 10.3390/su11215873.
- [118] J. I. Peláez, F. E. Cabrera, and L. G. Vargas, “Estimating the importance of consumer purchasing criteria in digital ecosystems,” *Knowledge-Based Systems*, vol. 162, no. July, pp. 252–264, 2018, doi: 10.1016/j.knosys.2018.07.023.
- [119] G. Vaccaro, F. E. Cabrera, J. I. Pelaez, and L. G. Vargas, “Comparison matrix geometric index: A qualitative online reputation metric,” *Applied Soft Computing*, vol. 96, p. 106687, Nov. 2020, doi: 10.1016/j.asoc.2020.106687.
- [120] J. I. Pelaez, A. Casado, E. R. Yanez, F. E. Cabrera, and A. H. Yanez, “A model for social assessment of governmental areas through intangible assets,” *2017 4th International Conference on eDemocracy and eGovernment, ICEDEG 2017*, pp. 200–204, 2017, doi: 10.1109/ICEDEG.2017.7962534.