

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS E INGENIERÍA**



**MARCO TEORICO PARA DIAGNÓSTICO Y PROPUESTA DE  
MEJORA EN LA RED DE AGENCIAS DE UNA EMPRESA BANCARIA  
DEL RUBRO DE LAS MICROFINANZAS EN LIMA  
METROPOLITANA EMPLEANDO HERRAMIENTAS DE  
MARKETING ANALYTICS Y DATA SCIENCE**

**Trabajo de investigación para la obtención del grado de BACHILLER EN  
CIENCIAS CON MENCIÓN EN INGENIERÍA INDUSTRIAL**

**AUTOR:**

Frank Esteban Sánchez Chicana

**ASESOR:**

Eduardo Carbajal López

Lima, Agosto, 2020

## Resumen

El trabajo de investigación tiene la finalidad de brindar el soporte teórico adecuado para la elaboración del trabajo de tesis homónimo, dotando de las herramientas de diagnóstico, para indagar sobre la estructura con la que se gobierna la empresa y encontrar una contramedida útil contra la causa raíz de la problemática de la organización, y de mejora, para planear una propuesta de mejora que sea factible técnica y económicamente a través de la aplicación de herramientas de *Marketing Analytics* y *Data Science*. Asimismo, se presentan tres casos de estudio que ofrecen soluciones similares en un entorno laboral parecido al de la entidad donde se aplicará el trabajo de tesis. De estas experiencias, se obtiene un punto de inicio sobre la manera de abordar un problema de analítica de datos e implementar una solución adecuada. Los resultados se enmarcan en la recolección de información necesaria para el desarrollo del trabajo académico, además de la decisión de tomar el *clustering* como herramienta principal al elaborar la propuesta de mejora sobre la empresa en cuestión. El documento concluye con la exposición de las técnicas recopiladas en el desarrollo de este y los aprendizajes obtenidos de cada caso de estudio que, de manera resumida, interceden en el campo de la RFM, el *clustering*, las redes neuronales y el análisis de sentimientos.

## Tabla de Contenidos

Resumen.....	i
Tabla de Contenidos .....	ii
Índice de Tablas .....	iv
Índice de Figuras.....	v
1. Herramientas de Diagnóstico.....	1
1.1. Mapeo de Macroprocesos .....	1
1.2. Matriz de comparaciones pareadas .....	4
1.3. Matriz QFD.....	6
1.4. Matriz de Priorización.....	8
1.5. Flujograma .....	9
1.6. Fichas Indicador.....	12
1.7. Diagrama de Pareto.....	13
1.8. Diagrama de Ishikawa.....	14
1.9. <i>Five Whys</i> .....	16
1.10. Matriz FACTIS .....	16
2. Herramientas del Modelo de Mejora .....	19
2.1. Conceptos Útiles de <i>Marketing Analytics</i> y <i>Data Science</i> .....	19
2.2. Regresión Lineal .....	21
2.3. Regresión Logística .....	22
2.4. <i>Clustering</i> .....	23
2.5. <i>Software</i> .....	25
3. Estudio de Casos.....	26
3.1. Primer Caso.....	26
3.2. Segundo Caso.....	30

3.3. Tercer Caso .....	36
4. Resumen Metodológico .....	43
Conclusiones .....	45
Bibliografía .....	46



## Índice de Tablas

Tabla 1 Escala de preferencias.....	4
Tabla 2 Matriz de comparaciones pareadas para la elección de un software .....	6
Tabla 3 Plantilla de una matriz de priorización .....	9
Tabla 4 Matriz de priorización sobre seguridad ciudadana .....	10
Tabla 5 Simbología ISO (parte 1).....	11
Tabla 6 Simbología ISO (parte 2).....	11
Tabla 7 Causas de reclamación costeadas .....	14
Tabla 8 Criterios y factores de ponderación FACTIS .....	17
Tabla 9 Ranking de factores para contramedidas .....	18
Tabla 10 <i>Clustering Performance</i> .....	29
Tabla 11 Características del clúster .....	30
Tabla 12 Valores del índice de evaluación de validez de las distintas cantidades de <i>cluster</i> ..	38
Tabla 13 Cantidad de <i>cluster</i> de las sucursales basada en los métodos de <i>clustering</i> .....	39
Tabla 14 Criterios de error para los resultados de los cálculos.....	42

## Índice de Figuras

Figura 1 Jerarquía de procesos.....	2
Figura 2 Mapa de procesos .....	3
Figura 3 Mapa de procesos de un hotel .....	4
Figura 4 La casa de la calidad.....	7
Figura 5 Matriz QFD del sacacorchos Mk11 .....	8
Figura 6 Ejemplo de diagrama de flujo con la simbología de la norma ISO.....	12
Figura 7 Ejemplo de ficha de indicador.....	13
Figura 8 Ejemplo de diagrama de Pareto .....	15
Figura 9 Ejemplo del diagrama de Ishikawa.....	15
Figura 10 <i>Five Whys</i> aplicado a una máquina detenida.....	16
Figura 11 Clasificación usando un <i>dataset</i> predefinido.....	22
Figura 12 Dendogramas con diferente cantidad de clústeres.....	24
Figura 13 Marco de trabajo integrado propuesto sobre el SMA.....	35
Figura 14 Valores de la distribución del logaritmo posterior en base a la cantidad de <i>cluster</i> .....	39
Figura 15 Diferencia entre el flujo real de entrada y salida de dinero de 18 agencias durante 5 meses comparada con la estimación por el método bayesiano de redes neuronales considerando los resultados obtenidos del <i>clustering</i> clásico y el bayesiano .....	40
Figura 16 Diferencia entre el flujo real de entrada y salida de dinero de 18 agencias durante 5 meses comparada con la estimación por el método clásico de redes neuronales considerando los resultados obtenidos del <i>clustering</i> clásico y el bayesiano .....	41

## 1. Herramientas de Diagnóstico

El primer punto del capítulo tiene el objetivo de establecer la guía de las herramientas que se utilizan en la investigación para realizar el diagnóstico de la empresa e identificar la causas raíz que conllevarán a la elección de la contramedida de mayor valor para los fines del proyecto.

### 1.1. Mapeo de Macroprocesos

Un proceso, según Cadena, es un conglomerado de actividades que disponen de una diversidad de entradas provistas por agentes (proveedores) externos e internos a la organización. Estos elementos son, a su vez, transformados y dotados de valor agregado con la finalidad de brindar un producto y/o servicio (2016: 62).

Cadena continúa su explicación asegurando que cualquier proceso posee la siguiente serie de componentes.

1. Entradas (inputs). Elementos otorgados por un proveedor que ingresan al sistema para ser transformados.
2. Mecanismos o recursos. Medios a través de los cuales se lleva a cabo el proceso.
3. Salidas (outputs). Productos y servicios entregados al cliente una vez que el proceso ha culminado.
4. Controles. Reglamentos, normas, leyes, procedimientos y políticas que se encargan de regular el desempeño y la ejecución del proceso.
5. Límites del proceso. Los puntos que definen el inicio y el fin del proceso.

Seguidamente, el autor clasifica los procesos en tres agrupaciones cuyo detalle se menciona en las siguientes líneas.

- Procesos estratégicos. Definen las bases que guían a los procesos misionales y de soporte con el objetivo de ser consecuentes con la visión de la organización.

- Procesos misionales. Se encargan de la generación de los productos y/o servicios ofertados por la organización en relación con su misión y objetivos.
- Procesos de soporte. Otorgan apoyo y sostén a los procesos misionales que lo requieran de acuerdo con sus actividades.

Por otro lado, Cadena desarrolla la jerarquía que gobierna sobre los subconjuntos y actividades que integran cada proceso. Por ello, como se muestra en la Figura 1, se establece una descomposición ordenada de un primer elemento llamado macroproceso, del que se desprenden los procesos que a su vez se componen de subprocessos, quienes están integrados por actividades.

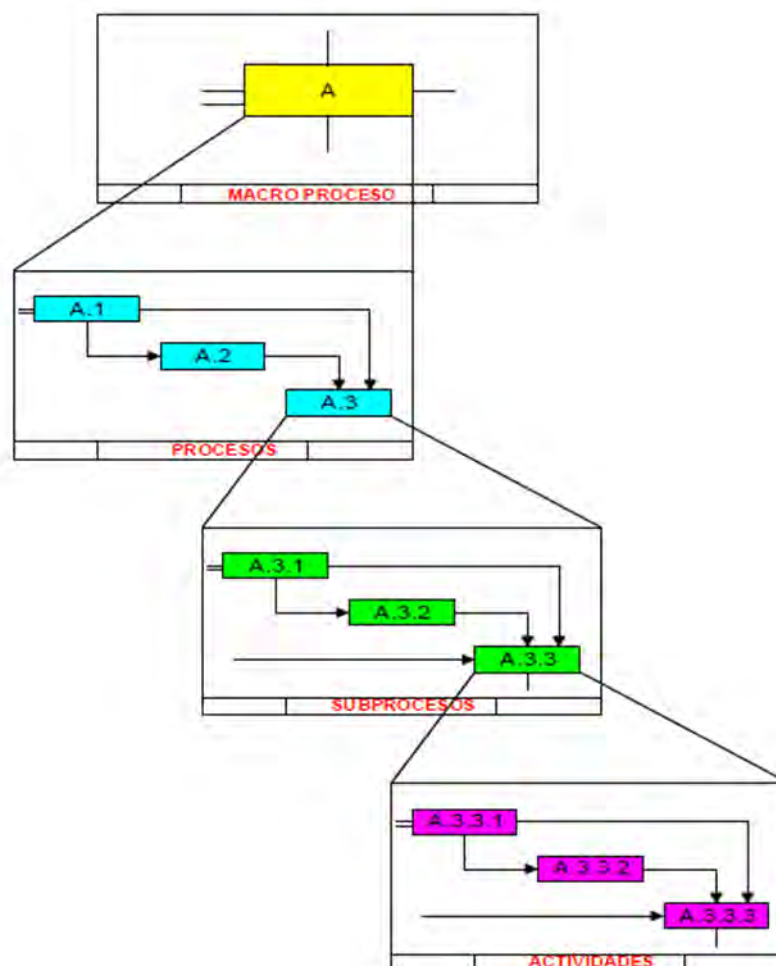


Figura 1 Jerarquía de procesos

Tomado de “Guía para el diseño y documentación de procesos”, por Cadena, 2016.

A partir de lo explicado, en la Figura 2 se presenta la composición básica de un mapa de procesos —también aplicable a una serie de macroprocesos—, integrado por los procesos estratégicos (gobernantes), misionales (productivos) y de soporte (de apoyo). Cada agrupación se presenta en un escaño diferenciado. Asimismo, se ubican dos pilares a los extremos izquierdo y derecho del diagrama que representan a los clientes. Estos últimos actores presentan sus necesidades y expectativas a modo de entrada y reciben un determinado nivel de satisfacción a la salida del conglomerado de procesos.

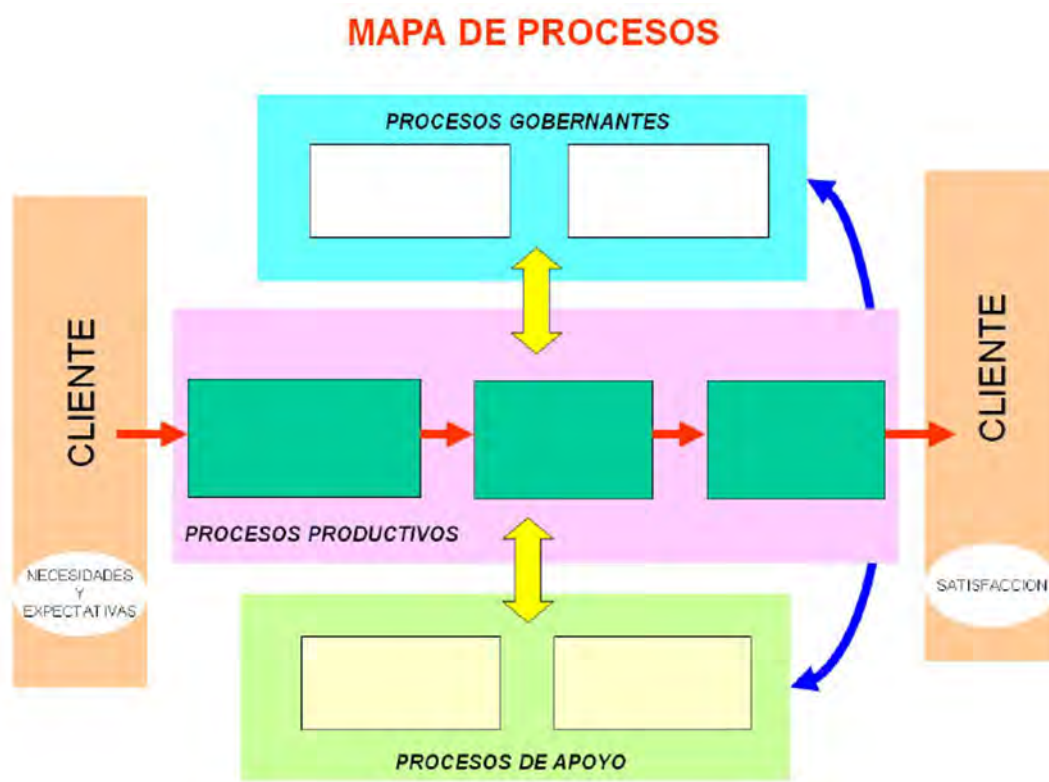


Figura 2 Mapa de procesos

Tomado de "Guía para el diseño y documentación de procesos", por Cadena, 2016.

A modo de ejemplo, se muestra la Figura 3 donde se plantea el mapa de procesos de un hotel. Los procesos estratégicos están conformados por el sistema de dirección, planificación estratégica y el sistema de calidad. Los procesos misionales siguen la ruta básica del ejercicio de un hotel que comprende la comercialización, reservas, recepción, estancia, restauración, eventos, facturación y despedida. Es en este punto donde se da la mayor interacción con el cliente y el servicio del hotel se hace tangible. Por último, están los procesos de soporte como

el mantenimiento, gestión de proveedores, servicios de administración y gestión de las personas. Para los fines del negocio, estos aseguran la disponibilidad de recursos y ambientes, así como la gestión holística de la organización en tiempo real.

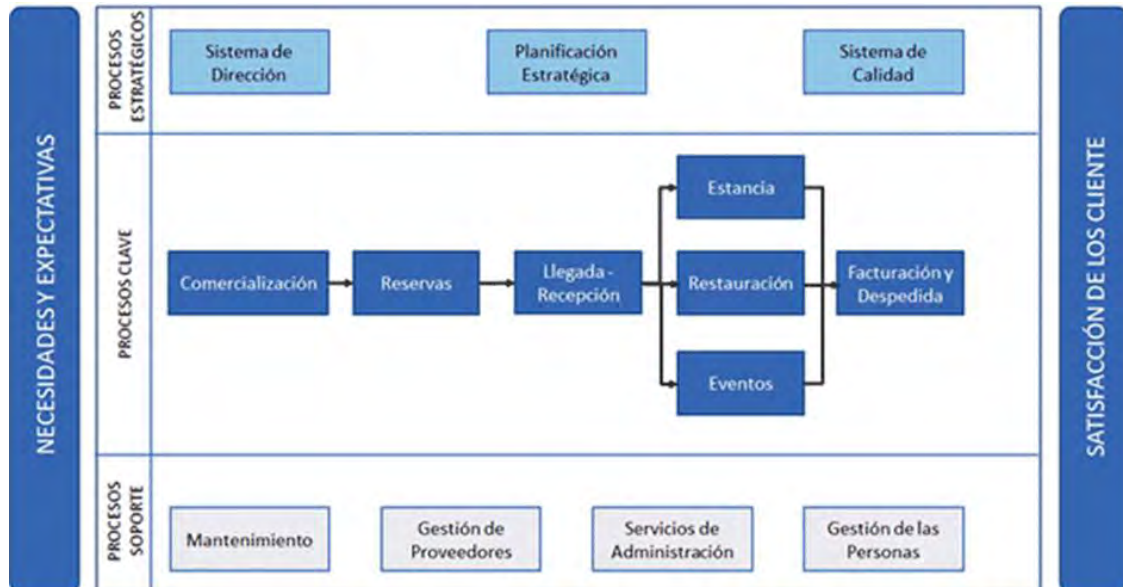


Figura 3 Mapa de procesos de un hotel

Tomado de “*Mapa de Procesos. Gestionar los Procesos*”, por Aiteco Consultores, s.f.

## 1.2. Matriz de comparaciones pareadas

Las comparaciones pareadas se elaboran sobre el contraste de la influencia de dos elementos bajo evaluación. Esta interacción es representada a través de un valor numérico entre 1 y 9 que vislumbra la preferencia por una de las alternativas sobre la otra. Esta escala es presentada en la Tabla 1 junto al planteamiento verbal que cada valor simboliza.

Tabla 1 Escala de preferencias

Planteamiento verbal de la preferencia	Calificación Numérica
Extremadamente preferible	9
Entre muy fuertemente y extremadamente preferible	8
Muy fuertemente preferible	7
Entre fuertemente y muy fuertemente preferible	6
Fuertemente preferible	5
Entre moderadamente y fuertemente preferible	4
Moderadamente preferible	3
Entre igualmente y moderadamente preferible	2
Igualmente preferible	1

Tomado de “*El Proceso de análisis jerárquico (AHP) como herramienta para la toma de decisiones en la selección de proveedores: aplicación en la selección del proveedor para la Empresa Gráfica Comercial MyE S.R.L.*”, por Toskano, 2005.

La matriz de comparaciones pareadas parte de una base matemática, tal y como lo define Toskano en su obra. “Sea  $A$  una matriz  $n \times n$ , donde  $n \in \mathbb{Z}^+$ . Sea  $a_{ij}$  el elemento  $(i, j)$  de  $A$  para  $i = 1, 2, \dots, n$ , y,  $j = 1, 2, \dots, n$ . Decimos que  $A$  es una matriz de comparaciones pareadas de  $n$  alternativas, si  $a_{ij}$  es la medida de la preferencia de la alternativa en el reglón  $i$  cuando se le compara con la alternativa de la columna  $j$ . Cuando  $i = j$ , el valor de  $a_{ij}$  será igual a 1, pues se está comparando la alternativa consigo misma” (2005: 26).

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & 1 \end{pmatrix}$$

Además, se cumple que:  $a_{ij} \cdot a_{ji} = 1$ ; es decir:

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{pmatrix}$$

De esta manera, los valores que toma cada elemento  $a_{ij}$  son los definidos de acuerdo con la escala de preferencia sobre cada criterio. Por lo tanto, la matriz de comparaciones pareadas toma la siguiente forma.

	Criterio 1	Criterio 2	...	Criterio $m$
Alternativa 1	$P_{11}$	$P_{12}$	...	$P_{1m}$
Alternativa 2	$P_{21}$	$P_{22}$	...	$P_{2m}$
...	...	...	...	...
Alternativa $n$	$P_{n1}$	$P_{n2}$	...	$P_{nm}$

La Tabla 2 expone un ejemplo de este tipo de matriz. En él se observan cinco criterios que son comparados uno a uno –con la excepción de que en esta ocasión usan una escala del 1 al 5 para asignar el grado de preferencia- sobre la diagonal superior. Estos valores son

reflejados en la diagonal inferior con su expresión inversa para luego calcular el peso ponderado definido de acuerdo con la interacción del conjunto de criterios entre sí.

Tabla 2 Matriz de comparaciones pareadas para la elección de un software

	Facilidad de uso	Acceso e integración de fuentes de datos	Creación de visualización y paneles de datos	Soporte para R y Python	Precio	Total	Peso ponderado definido
Facilidad de uso		5	1	5	5	16,00	0,44
Acceso e integración de fuentes de datos	0,2		1	5	1	7,20	0,20
Creación de visualización y paneles de datos	1	1		1	5	8,00	0,22
Soporte para R y Python	0,2	0,2	1		1	2,40	0,07
Precio	0,2	1	0,2	1		2,40	0,07
Total						36	1,00

Tomado de “Cómo hacer una matriz de priorización”, por Betancourt, 2018.

### 1.3. Matriz QFD

De acuerdo con Bernal, la matriz QFD es “un sistema que tiene como objetivo interpretar e incorporar la ‘Voz del Consumidor’ en las características de calidad de productos, procesos y servicios que permitan alcanzar la satisfacción de las necesidades [del] cliente” (2009: 3).

La HOQ (Casa de la Calidad por sus siglas en inglés) es una matriz que utiliza la metodología QFD para su elaboración. En la Figura 4 se observa la estructura que compone este tipo de diagrama cuya similitud con una casa es evidente. Seguidamente, se detalla a grandes rasgos la definición de cada elemento del esquema.

1. *Voice of the Costumer* (VOC). Es el listado de requerimientos del cliente que evoca a los objetivos de la metodología, el “qué” es deseado para el desarrollo del servicio.
2. Requerimientos de diseño. Corresponden a las características o criterios bajo los cuales se asigna la puntuación de cada elemento al finalizar el desarrollo. Asimismo, responden al “cómo” será satisfecho cada requerimiento.

3. Matriz de relaciones. Es capaz de relacionar los “qué” y los “cómo” a través de tres valores: 1, 3 y 9, correspondientes a una débil, mediana y fuerte relación. Una casilla en blanco denota la ausencia de influencia de forma relevante.
4. *Benchmarking*. Permite valorar el desempeño de los competidores bajo los mismos conceptos de “qué” y “cómo” y compararlo con el ofrecido.
5. Ponderación. Determina la relevancia de los “qué” y los “cómo” frente al cumplimiento de la meta propuesta. Los primeros se priorizan de acuerdo con la evaluación del cliente usando una escala de 1 a 5 o de 1 a 10.
6. Matriz de correlación. Permite determinar la intensidad de las correlaciones entre los requerimientos de diseño con el fin de identificar relaciones de apoyo o conflicto. Los valores pueden ser: altamente positiva (++), positiva (+), negativa (-) y altamente negativa (--).

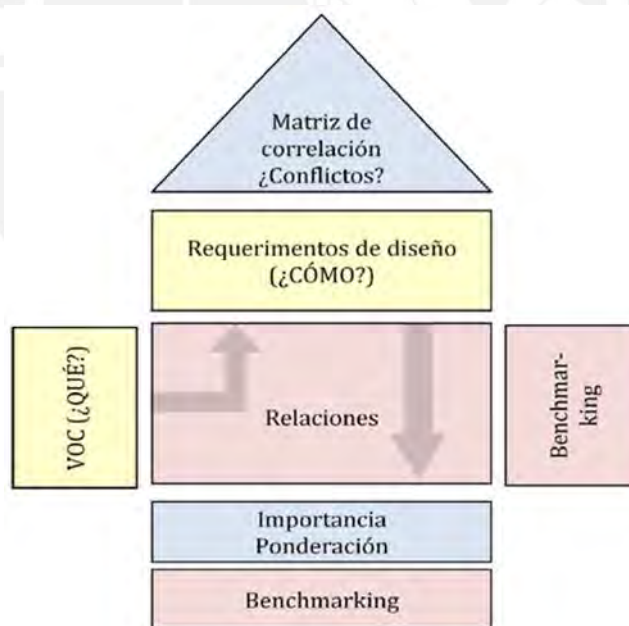


Figura 4 La casa de la calidad

Tomado de “*Quality Function Deployment (QFD) para servicio*”, por Bernal, Dornberger y Suvelza, 2009.

La Figura 5 presenta un ejemplo donde se utiliza la herramienta de la Casa de la Calidad en el desarrollo de un sacacorchos. Los “qué” está agrupados en cuatro subgrupos: seguridad, diseño, desempeño y calidad. Con respecto a los “cómo”, se ha considerado una lista de

critérios técnicos sobre la funcionalidad, capacidad y mecanismo del artículo. Asimismo, el apartado de Benchmarking utiliza como competidores a otros modelos de sacacorchos. Luego de aplicar los puntajes acordes al criterio del autor, se define que el requerimiento de diseño con mayor relevancia es el torque aplicado durante el uso del artículo.

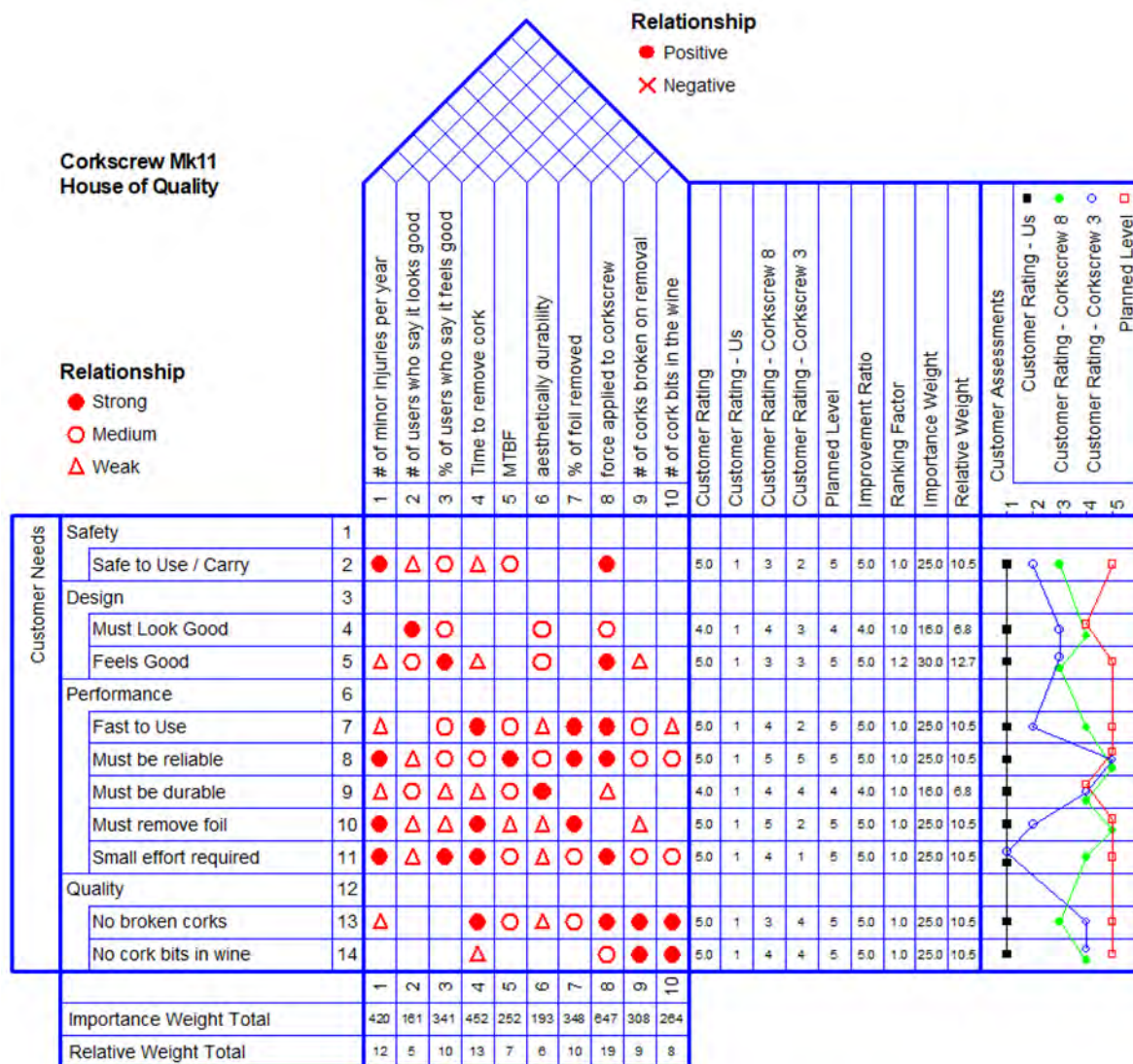


Figura 5 Matriz QFD del sacacorchos Mk11

Tomado de “Quality Function Deployment - QFD2000 Pro”, por Arcturus7, 2019.

### 1.4. Matriz de Priorización

La matriz de priorización es una herramienta útil cuando se trata de comparar y elegir de forma racional entre una serie de alternativas bajo ciertos criterios con el objetivo de determinar la prioridad de cada una (Leebov 1993).

En la Tabla 3 se distingue el modelo básico de una matriz de priorización. En ella se expone que los criterios se ubican en las columnas de la tabla, mientras que las opciones se encuentran en las filas. No obstante, este orden puede cambiar, es decir, criterios y opciones pueden estar en las filas o columnas indistintamente. Asimismo, es posible utilizar un factor de ponderación por cada criterio con el fin de entregarle a cada uno un nivel de importancia diferenciado que devuelva un valor ponderado al calificar cada opción.

Tabla 3 Plantilla de una matriz de priorización

	Criterio 1	Criterio 2	Criterio 3	Criterio 4	Total
OPCION 1					
OPCION 2					
OPCION 3					
OPCION 4					
OPCION 5					
OPCION 6					
OPCION 7					

Tomado de “Manual de los administradores de salud para el mejoramiento continuo”, por Leebov, 1993.

La Tabla 4 muestra un ejemplo de matriz de priorización donde los elementos evaluados son los servicios listados en la fila superior de la tabla. En la primera columna se encuentran los criterios de evaluación, los cuales han sido ponderados con la ayuda de una matriz de comparaciones pareadas aplicada previamente. A continuación, se puntuaron los elementos sobre los criterios y se obtuvo un puntaje ponderado categorizado como “suma” al final de la tabla.

### 1.5. Flujograma

Un flujograma se define como “una mezcla de símbolos y explicaciones que expresan secuencialmente los pasos de un proceso, de forma tal que este se comprenda más fácilmente

[...]. En pocas palabras son la representación simbólica de los procedimientos administrativos” (Ministerio de Planificación Nacional y Política Económica 2009).

Tabla 4 Matriz de priorización sobre seguridad ciudadana

CRITERIOS (Objetivos Estratégicos)	POND.	SERVICIOS			
		Detención de Delincuentes e Infractores	Atención en Comisarias	Atención de Emergencias en Centros Médicos	Comunicación en caso de delitos y faltas
Promover la participación de los ciudadanos para enfrentar la inseguridad ciudadana	15%	4	1	1	3
Profesionalización de los recursos humanos de la PNP y serenazgos	35%	4	4	1	1
Planificación conjunta de la Policía y serenazgos para mejorar la vigilancia de los calles	39%	5	5	2	4
Mejorar el sistema de administración para la reducción de la delincuencia	11%	3	4	4	4
<b>SUMA</b>		<b>4.3</b>	<b>3.9</b>	<b>1.7</b>	<b>2.8</b>

Tomado de “*Diagnóstico y mejora para el servicio de la seguridad ciudadana en el distrito de San Martín de Porres*”, por Sotelo, 2017.

En su obra, el Ministerio de Planificación Nacional y Política Económica continúa con la presentación de las normas ISO (*International Organization for Standardization*), en especial la ISO 9000 que brinda los lineamientos gráficos para la elaboración de un diagrama de flujo orientado a la gestión continua de la calidad aplicable a cualquier organización que produce bienes o servicios. La simbología propuesta se muestra en la Tabla 5 y la Tabla 6.

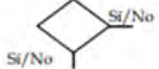




La Figura 6 muestra el ejemplo de un diagrama de flujo utilizando la simbología ISO. El mismo representa la ejecución de una maquila de ropa. Se observan las operaciones de producción y empaquetado, una inspección, los símbolos de entrada de bienes y almacenamiento, un elemento de operación e inspección y un símbolo de decisión.

Tabla 5 Simbología ISO (parte 1)

Símbolo	Significado	¿Para que se utiliza?
	<b>Operación</b>	Indica las principales fases del proceso, método o procedimiento.
	<b>Operación e Inspección</b>	Indica la verificación o supervisión durante las fases del proceso, método o procedimiento de sus componentes.
	<b>Inspección y Medición</b>	Representa el hecho de verificar la naturaleza, cantidad y calidad de los insumos y productos.
	<b>Transporte</b>	Indica cada vez que un documento se mueve o traslada a otra oficina y/o funcionario.
	<b>Entrada de bienes</b>	Indica productos o materiales que ingresan al proceso.
	<b>Almacenamiento</b>	Indica el depósito permanente de un documento o información dentro de un archivo.

Tomado de “*Guía para la elaboración de diagramas de flujo*”, por Ministerio de Planificación Nacional y Política Económica, 2009.

Tabla 6 Simbología ISO (parte 2)

Símbolo	Significado	¿Para que se utiliza?
	<b>Decisión</b>	Indica un punto dentro del flujo en que son posibles varios caminos alternativos.
	<b>Líneas de flujo</b>	Conecta los símbolos señalando el orden en que se deben realizar las distintas operaciones.
	<b>Demora</b>	Indica cuando un documento o el proceso se encuentra detenido, ya que se requiere la ejecución de otra operación o el tiempo de respuesta es lento.
	<b>Conector</b>	Conector dentro de página. Representa la continuidad del diagrama dentro de la misma página. Enlaza dos pasos no consecutivos en una misma página.
	<b>Conector de página</b>	Representa la continuidad del diagrama en otra página. Representa una conexión o enlace con otra hoja diferente en la que continua el diagrama de flujo.

Tomado de “*Guía para la elaboración de diagramas de flujo*”, por Ministerio de Planificación Nacional y Política Económica, 2009.

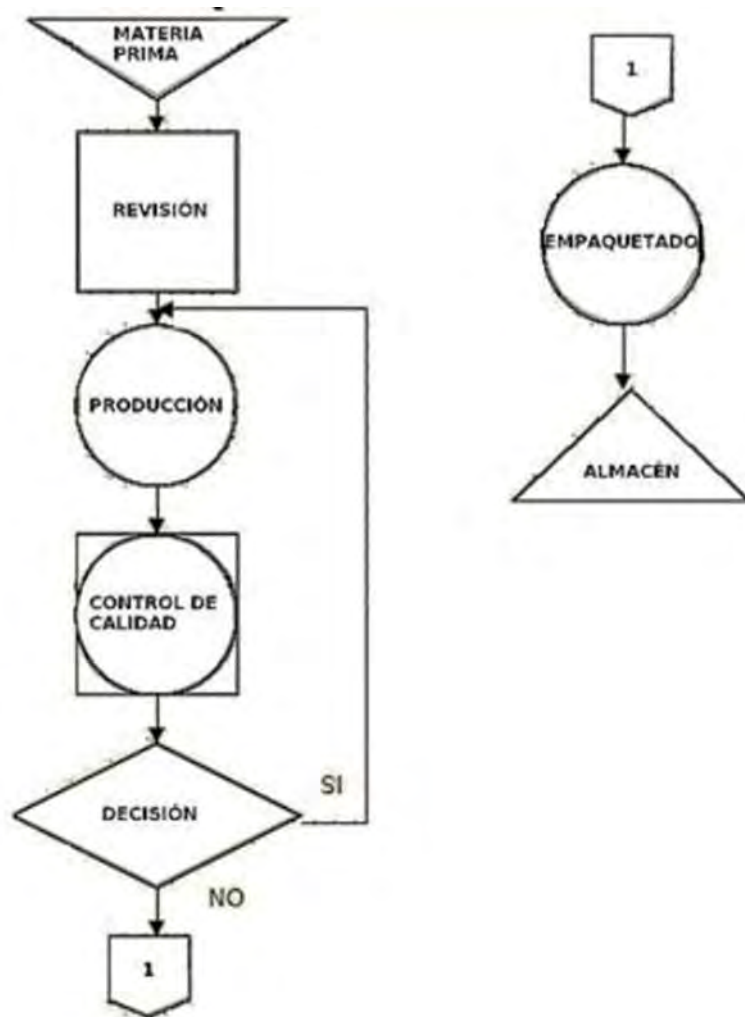


Figura 6 Ejemplo de diagrama de flujo con la simbología de la norma ISO

Tomado de “*Guía para la elaboración de diagramas de flujo*”, por Ministerio de Planificación Nacional y Política Económica, 2009.

### 1.6. Fichas Indicador

De acuerdo con Donneys, la ficha de indicador o ficha técnica es un formato que formaliza el indicador y el resultado alrededor de él. A través de este se puede evaluar las principales características del indicador para elaborar conclusiones sobre el sistema evaluado. Asimismo, la principal ventaja de la herramienta es la capacidad de analizar varios elementos a la vez (2009).

La Figura 7 muestra una ficha de indicador referida al costo de atención por materia de soporte técnico de productos de garantía. Entre los datos asignados al formato se encuentran la fórmula, el responsable, el tipo, la unidad, la fuente, la frecuencia, oportunidad, el glosario, los

datos y la representación gráfica del indicador donde se delimita mediante distintos colores los rangos de alcance de la meta propuesta.

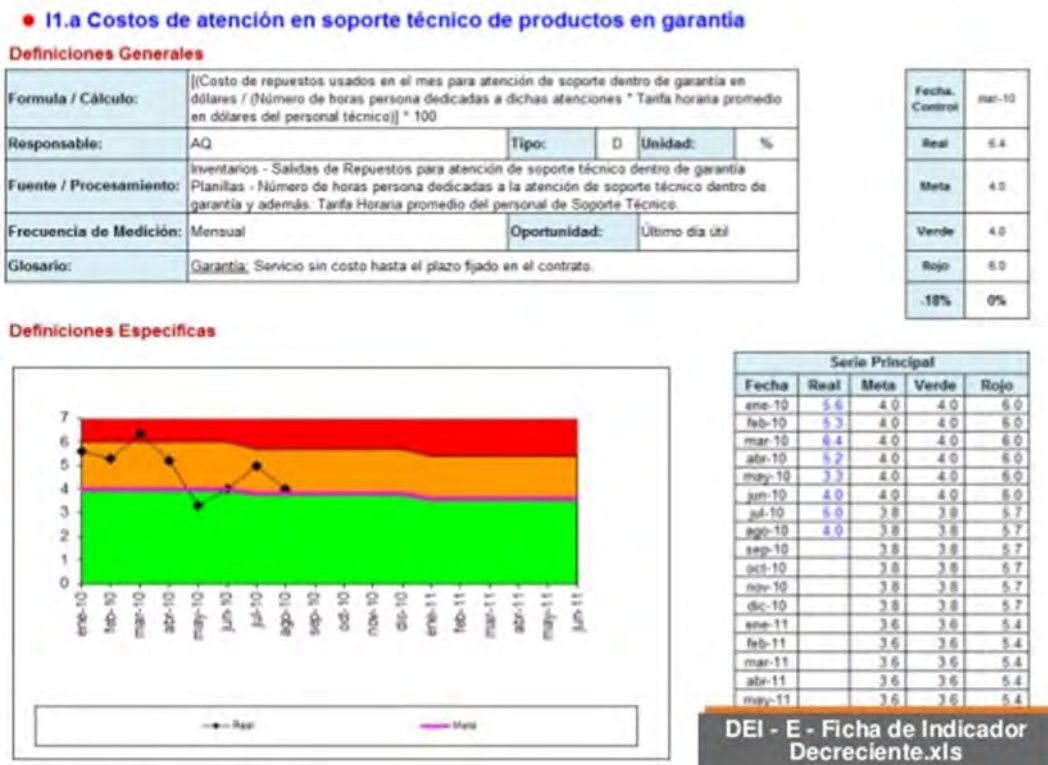


Figura 7 Ejemplo de ficha de indicador

Tomado de “DEI - C - Ficha de Indicador - 2012 11 25”, por Metodologías Más Soluciones, 2011.

### 1.7. Diagrama de Pareto

El diagrama de Pareto es la representación gráfica de la “ley 80-20” o de “los pocos vitales y muchos triviales”, la cual explica que al dividir las causas que originan un problema en alguna organización y cuantificarlas, tan solo unas pocas de ellas explican la mayor parte del efecto. De esta manera, la empresa es capaz de dirigir sus esfuerzos sobre esas causas ahora priorizadas. En resumen, en 20% de los factores se concentra un 80% del efecto (Escuela Técnica Superior de Ingenieros Industriales 2006).

En las siguientes líneas se presenta un ejemplo de aplicación del diagrama de Pareto donde la intención es identificar el motivo de los reclamos por parte de los clientes sobre los productos de una empresa del sector de las artes gráficas. En la Tabla 7 se muestra la lista de

causas de reclamos con mayor coste acompañados de su peso correspondiente frente al problema global.

Tabla 7 Causas de reclamación costeadas

Causa de la reclamación	Nº reclams.	Coste (miles pts.)	Porcentaje individual	Porcentaje acumulado
Defectos de color	5	700	37,4%	37,4%
Errores en el texto	15	530	28,3%	65,7%
Motas	13	150	8,0%	73,7%
Errores de corte	4	134	7,2%	80,9%
Mal troquelado	1	99	5,3%	86,2%
Calidad del papel	2	50	2,7%	88,9%
Máculas en trabajo final	7	49	2,6%	91,5%
Mala encuadernación	4	43	2,3%	93,8%
Daños en transporte	1	34	1,8%	95,6%
Entrega retrasada	22	25	1,3%	96,9%
Errores facturación	6	6	0,3%	97,2%
Otros	9	52,5	2,8%	100%
TOTAL	89	1.872,5		

Tomado de “*El diagrama de Pareto*”, por Escuela Técnica Superior de Ingenieros Industriales, 2006.

Seguidamente, todos los elementos son graficados en la Figura 8 de mayor a menor grado de importancia y señalados con su porcentaje acumulado con el fin de hallar la cantidad de elementos que representan ese 20% de las causas generando el 80% de los costos. De ella se identifica que aproximadamente los cuatro primeros elementos explican alrededor del 80% del costo.

### 1.8. Diagrama de Ishikawa

El diagrama de Ishikawa o diagrama causa-efecto es una herramienta que permite visualizar en un mismo plano de forma detallada las diferentes causas que pueden ocasionar un problema en particular. Para su elaboración, se recurre a la opinión de personas directa o indirectamente relacionadas con la problemática que puedan ofrecer datos importantes sobre su origen (Escuela Técnica Superior de Ingenieros Industriales 2006).

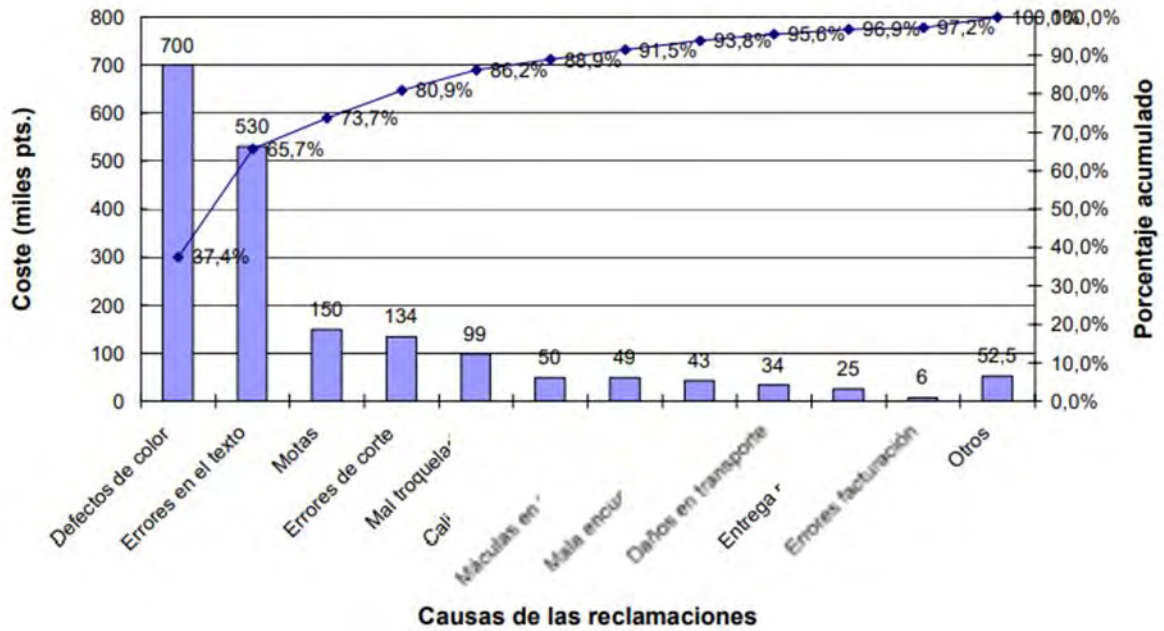


Figura 8 Ejemplo de diagrama de Pareto

Tomado de “El diagrama de Pareto”, por Escuela Técnica Superior de Ingenieros Industriales, 2006.

La Figura 9 exhibe un ejemplo de aplicación del diagrama de Ishikawa referido a la identificación de causas del desconocimiento de la marca en el mercado. Los factores han sido agrupados en seis categorías definidas como proveedores, recursos, mercado, personal, condiciones cambiantes del mercado y publicidad. Esta distribución permite la ubicación rápida de las causas relacionadas a un mismo tópico, así como alcanzar un orden.



Figura 9 Ejemplo del diagrama de Ishikawa

Tomado de “Diagrama de Ishikawa”, por Milián, s.f.

### 1.9. Five Whys

Los cinco ¿por qué? es una herramienta utilizada para identificar posibles causas raíz de un problema. La intención de su aplicación es buscar con suficiente profundidad aquellas causas escondidas en la superficie. Asimismo, esta técnica requiere que se realice la pregunta “¿por qué?” al menos cinco veces hasta que sea complicado pues se habrá encontrado aquellas causas raíz (Puga 2018).

La Figura 10 ofrece un ejemplo de aplicación de *Five Whys* donde el problema es una máquina detenida. Se desarrolla la pregunta cinco veces consecutivas identificando causa tras causa para el acontecimiento hasta llegar a la raíz del problema: no había filtro.

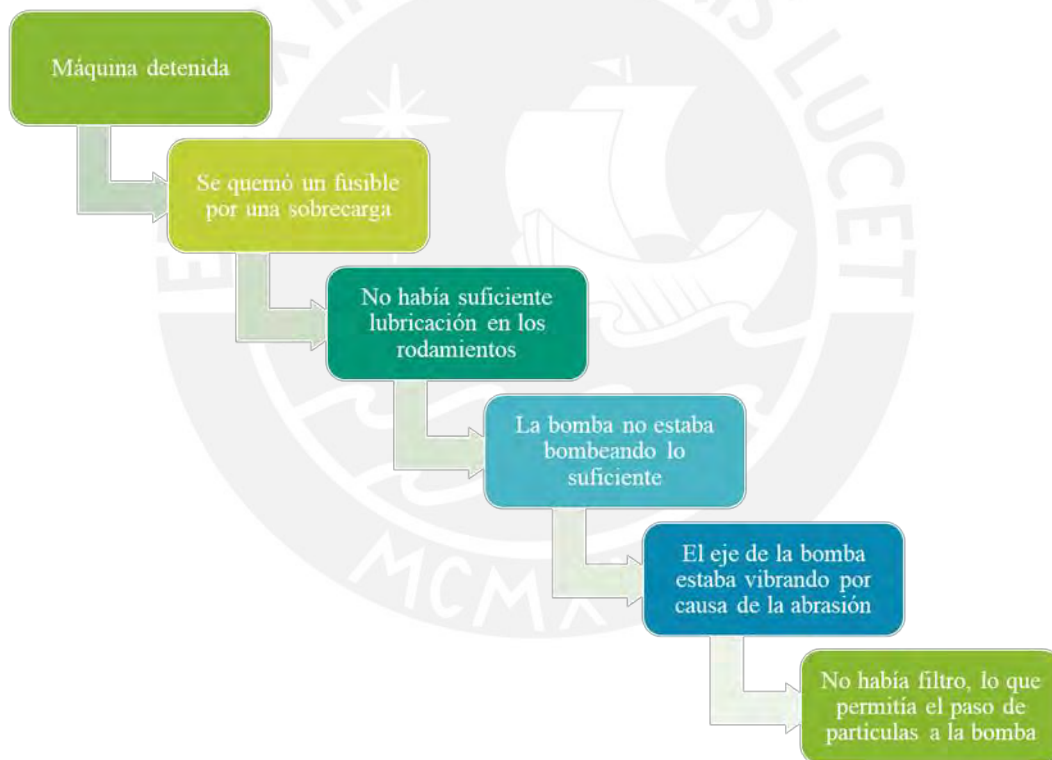


Figura 10 *Five Whys* aplicado a una máquina detenida

Tomado de “*Los cinco por qué (Five Whys)*”, por Puga, 2018.

### 1.10. Matriz FACTIS

La matriz FACTIS es una variante de la matriz de priorización que permite elegir una opción o solución frente a una problemática bajo una determinada coyuntura. Esta decisión es

tomada con base en valores cuantificables que han sido definidos junto a los criterios de evaluación, cada uno con un rango de tres posibles valores.

1. F. Denota la facilidad con la que la solución propuesta puede ser implementada en el sistema.
2. A. Indica si su despliegue genera un efecto sobre otras áreas de la organización.
3. C. Señala si se percibe una mejora en la calidad a partir del efecto de la propuesta.
4. T. Valora el tiempo que toma implementar cada solución.
5. I. Prioriza la inversión requerida por cada alternativa.
6. S. Establece cuál es el nivel de seguridad percibido en la actividad.

La Tabla 8 muestra el primer arreglo para la aplicación de FACTIS en la toma de una decisión. En esta primera vista, se observan los seis criterios descritos acompañados de una escala de tres valores que puede fluctuar de la forma 1-2-3 o 1-3-5. Los valores más altos representan un aspecto positivo para la elección de la alternativa. Asimismo, un factor de ponderación se cierne al lado derecho de la tabla para definir la importancia de cada categoría en la evaluación.

Tabla 8 Criterios y factores de ponderación FACTIS

CRITERIOS DE SELECCIÓN			FACTOR DE PONDERACIÓN	
F	Facilidad de implementación			1
	1: Muy difícil	2: Difícil	3: Fácil	
A	Afecta a otras áreas su implementación			5
	1: Sí	3: Medio	5: Nada	
C	Mejoramiento de la calidad			6
	1: Poco	3: Medio	5: Mucho	
T	Tiempo que implica implementarlo			4
	1: Largo plazo	2: Medio plazo	3: Corto plazo	
I	Inversión requerida			2
	1: Alta	3: Media	5: Poca	
S	Nivel de seguridad en el servicio			3
	1: Poco	3: Medio	5: Alta	

Tomado de “Diagnóstico y mejora para el servicio de la seguridad ciudadana en el distrito de San Martín de Porres”, por Sotelo, 2017.

En segundo lugar, se asignan los valores de la escala definida por cada criterio para cada alternativa en una matriz como en la Tabla 9. Seguidamente, se realizan los cálculos numéricos necesarios para obtener la suma ponderada del puntaje de cada quién y, con ello, se elige la alternativa más robusta y viable.

Tabla 9 Ranking de factores para contramedidas

	F	A	C	T	I	S	
CONTRAMEDIDAS	Facilidad para implementar contramedida	Implementación afecta a otras áreas	Mejoramiento de la calidad	Tiempo de Implementación	Inversión requerida	Seguridad	TOTAL
Mapa de distribución de recursos	3	5	5	2	3	5	87
Herramienta DMAIC	2	3	5	2	3	5	76
Programa de distribución de presupuesto	2	1	5	3	3	3	64
Capacitaciones y programas de soporte	3	3	3	3	5	5	73

Tomado de “Diagnóstico y mejora para el servicio de la seguridad ciudadana en el distrito de San Martín de Porres”, por Sotelo, 2017.

## 2. Herramientas del Modelo de Mejora

Con la contramedida seleccionada a partir del análisis de los conceptos relevantes con las herramientas de diagnóstico, se inicia el desarrollo de la propuesta de mejora. Para ello, se dispone de una serie de herramientas que específicamente se encargan de la elaboración y ejecución de la propuesta en el tercer capítulo. En este acápite, se explican estos conceptos como se lista a continuación.

### 2.1. Conceptos Útiles de *Marketing Analytics* y *Data Science*

Los conceptos explicados en este acápite corresponden a términos útiles e interesantes que permiten la comprensión del proceso a seguir durante el desarrollo de la propuesta de mejora.

#### 1. *Data mining*

La minería de datos se explica cómo el proceso de explorar información y encontrar patrones en ella usando *Machine Learning*, “aprendizaje de máquina” en español, estadística y bases de datos. El objetivo principal es obtener información útil a partir de los datos que pueda ser utilizada para incrementar las ganancias, reducir costos o incluso salvar vidas a través de su uso. Asimismo, es posible acelerar este proceso con el conocimiento de dominios, cuya importancia radica en comprender del significado de los datos y saber cómo analizarla para obtener propuestas de mejora (Madhavan 2015).

#### 2. *Machine Learning*

El aprendizaje de máquina es una técnica para entrenar a ciertos programas que usan data a generar algoritmos en lugar de programar cada uno explícitamente desde cero. Además, se puede definir como un campo de la ciencia computacional originada a partir de la investigación sobre inteligencia artificial. Se encuentra estrechamente relacionado a la estadística y la optimización matemática. Por último, algunos ejemplos aplicativos de esta técnica se observan

en filtros de spam electrónico, motores de búsqueda, traducción de idiomas, entre otros (Madhavan 2015: 107).

Madhavan continúa con el tópico acotando que, dependiendo de la naturaleza del aprendizaje planteado como objetivo o la retroalimentación disponible para el sistema, el aprendizaje de máquina se clasifica en tres tipos.

- Aprendizaje supervisado. La computadora recibe una cantidad determinada de inputs y sus respectivos outputs. El propósito del programa es el de aprender de los inputs para reproducir los outputs. Algunos ejemplos comprenden árboles de decisión, regresión lineal, regresión logística y clasificadores de Naive Bayes.
- Aprendizaje no supervisado. No existe una variable objetivo. La computadora trabaja por su cuenta para encontrar patrones dentro de la información. Algunos ejemplos comprenden la clusterización de k-medias y la jerárquica.
- Aprendizaje de reforzamiento. El programa debe interactuar con el entorno de forma dinámica, justo como conducir un auto. Algunos ejemplos comprenden el aprendizaje de diferencias temporales y *Q learning*.

### 3. *Marketing Analytics*

La analítica del *Marketing* contiene los procesos y tecnologías que permiten evaluar el éxito de las iniciativas de *Marketing*. Para ello, dispone de métricas de negocio relevantes para el mercado como el ROI para exponer el verdadero desempeño de un programa de *Marketing*. Asimismo, la capacidad de esta versión de la analítica se conjuga en tres aspectos: descriptiva, predictiva y prescriptiva (SAS)

- Analítica descriptiva. Reporta el pasado para responder incógnitas como qué elementos de la última campaña fueron más rentables, cómo contrastaron dos campañas de correo electrónico y cuántas iniciativas se generaron durante la última temporada.

- Analítica predictiva. Analiza el presente para responder cómo los consumidores están siendo capturados por la empresa, qué canales de comunicación prefieren y cuánto impacto tiene la marca en redes sociales.
- Analítica prescriptiva. Predice e influencia el futuro con las respuestas a cómo asegurar la lealtad de los clientes, cuán sensible sería la rentabilidad ante cambios en el salario y qué regiones estarán usando un servicio de la empresa en el futuro.

## 2.2. Regresión Lineal

La regresión lineal es útil para estimar respuestas cuantitativas. Es una herramienta útil cuando su funcionamiento es combinado con un método de aprendizaje estadístico. En las siguientes líneas se presenta la versión simple y la múltiple.

### 1. Regresión lineal simple

James (2013) explica que se trata de una aproximación sencilla a la predicción de una variable respuesta  $Y$  a partir de una variable predictora  $X$ . Esta asume que existe una relación aproximadamente lineal entre ambas variables. De forma matemática, se forma la siguiente expresión.

$$Y \approx \beta_0 + \beta_1 X$$

Si se comprende la relación entre  $X$  e  $Y$  bajo la forma  $Y = f(X) + \epsilon$  para alguna función  $f$  donde  $\epsilon$  es el término de error aleatorio de media cero. Aproximado a una función lineal, la relación entre ambas variables se puede modelar de la siguiente forma.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Donde  $\beta_0$  es el intercepto —es decir, el valor que toma  $Y$  cuando  $X = 0$ — y  $\beta_1$  es la mediana —es decir, el incremento promedio de  $Y$  debido al crecimiento de  $X$  en una unidad— y coeficiente de  $X$ .

### 2. Regresión lineal múltiple

James (2013) continúa con este tipo de regresión lineal indicando que elimina la necesidad de realizar un modelo de regresión lineal simple para cada predictor. En cambio, se extiende el modelo básico para ser influenciado por múltiples predictores. De esta manera, cada variable predictora es acompañada de un coeficiente. En general, se asume que existen  $p$  distintas variables predictoras. Por ello, el modelo de regresión múltiple toma la siguiente forma.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

### 2.3. Regresión Logística

Trabaja de forma distinta a una regresión lineal puesto que el valor entregado por una regresión logística solo varía entre 0 y 1 como una tendencia a la ocurrencia de un evento específico. En la Figura 11 se observa del lado izquierdo una regresión lineal simple estimando los valores futuros inconsistentemente debido a que resulta en algunos valores por debajo de cero. En cambio, el gráfico de la derecha muestra una regresión logística que estima valores que fluctúan entre 0 y 1 (James 2013: 131).

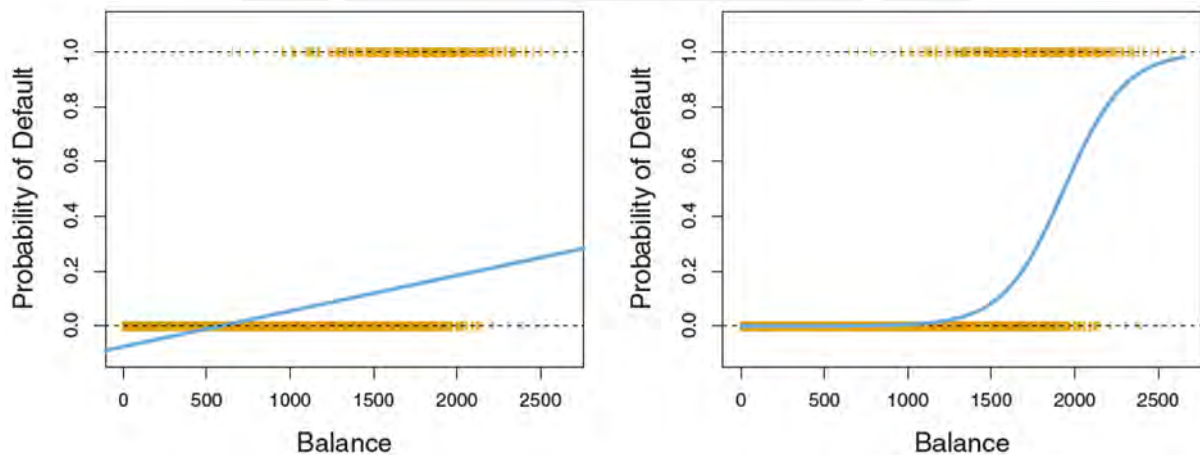


Figura 11 Clasificación usando un *dataset* predefinido

Tomado de “An Introduction to Statistical Learning with Applications in R”, por James, 2013.

La ecuación que expresa en su forma más básica el modelo de regresión logística se presenta a continuación. El significado de cada variable y coeficiente sigue la misma lógica que para la regresión lineal.

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Por último, se mencionan algunos escenarios donde la regresión logística puede ser aplicada.

- Obtener el grado de predisposición de un cliente en una tienda minorista de comprar un nuevo producto en lanzamiento.
- Obtener la probabilidad de que un transformador falle utilizando los datos del sensor asociado a él.
- Obtener la probabilidad de que un cliente haga clic en el anuncio de un sitio web basado en su comportamiento en Internet.

#### 2.4. Clustering

De acuerdo con James, la clusterización se refiere a un conjunto de técnicas utilizadas para encontrar subgrupos o conglomerados en un conjunto de datos. Los elementos son reunidos en distintas agrupaciones con la característica de que cada uno de los ítems de un grupo en particular posee características similares entre sí, que posteriormente serán definidas a cabalidad por el usuario (2013: 385). Asimismo, existen dos tipos de clusterización disponibles como se revisa a continuación.

##### 1. *K-means clustering*

Tiene el objetivo de particionar un conjunto de datos en K distintos conglomerados que no se superponen entre sí. Primero, se declara un número deseado de conglomerados K, luego el algoritmo asigna cada observación a alguno de los clústeres de acuerdo con las distancias delimitadas entre sí (James 2013: 386).

La siguiente expresión representa la búsqueda del algoritmo por establecer y repartir cada elemento en un clúster de tal manera que la distancia entre ese elemento y su centroide sea la mínima. Donde  $|C_k|$  representa el número de observaciones en el k-ésimo clúster.

$$\text{minimize}_{C_1, \dots, C_K} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\}$$

## 2. Hierarchical clustering

Brinda un enfoque alternativo que no requiere una predefinición del número de clúster K por el usuario. Este tipo de clusterización posee una ventaja sobre *K-means clustering*, tiene la capacidad de ofrecer una representación similar a un árbol de las observaciones llamado dendograma (James 2013: 390). En la Figura 12 se muestra un grupo de dendogramas. El primero establece un único clúster que engloba todas las observaciones, el subsiguiente genera dos clústeres y el último, tres clústeres.

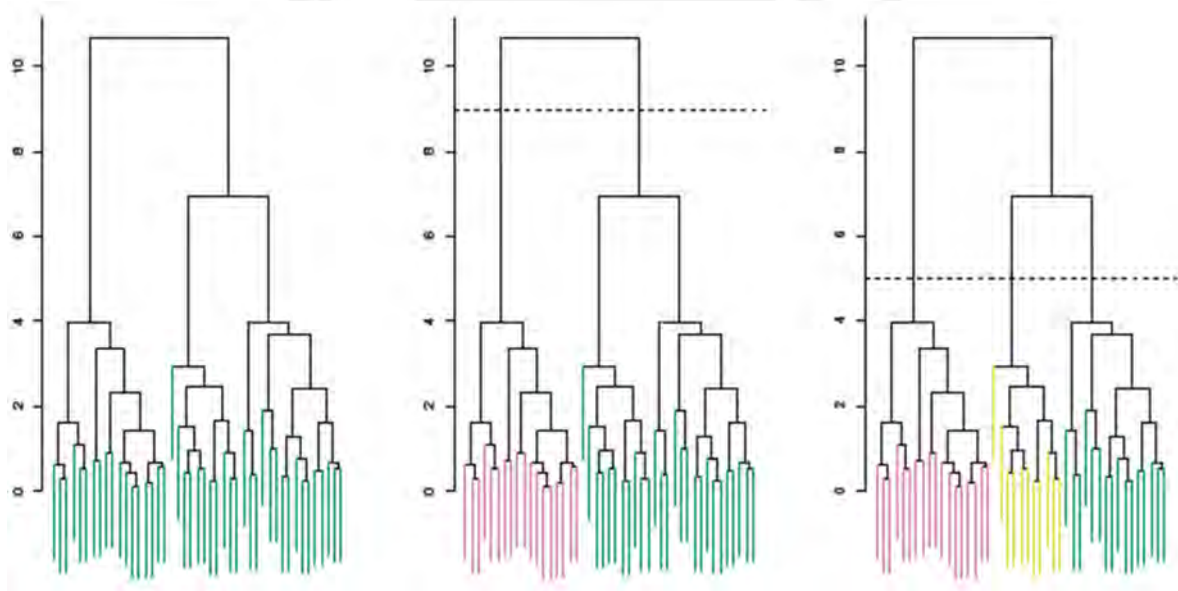


Figura 12 Dendogramas con diferente cantidad de clústeres

Tomado de “An Introduction to Statistical Learning with Applications in R”, por James, 2013.

El algoritmo para la conformación de los clústeres se explica a continuación.

- Se inicia con n observaciones y cada observación es tratada como su propio clúster.
- Se agrupan los elementos más próximos y forman un nuevo clúster dejando atrás los clústeres definidos previamente.
- Se itera de esta forma hasta consumir todos los elementos y obtener una agrupación de todas las observaciones iniciales.

## 2.5. *Software*

El software sobre el cual se apoya la propuesta de mejora es R. De acuerdo con su portal web oficial, es un lenguaje y entorno para realizar aplicaciones de estadística por computadora y gráficos. Este programa ofrece una amplia variedad de herramientas estadísticas y técnicas para graficar, así como una comunidad que continuamente innova y crea nuevos paquetes y librerías que permiten a los usuarios optimizar sus procesos de análisis dentro del *software*.

Además, ofrece a sus usuarios una serie de ventajas, entre las cuales se encuentran las listadas en las siguientes líneas.

- Un efectivo manejo y almacenamiento de datos
- Un entorno que permite realizar programación orientada a objetos
- Facilidad para graficar y visualizar los datos
- Un lenguaje de programación certificado, simple y efectivo

Los principales paquetes utilizados en el desarrollo de la propuesta de mejora son “lm” (regresión lineal), “glm” (regresión logística) y “kmeans” (clusterización).

### 3. Estudio de Casos

Con el objetivo de ganar una visión más amplia con respecto a la aplicación de las herramientas del modelo de mejora desarrolladas en el acápite anterior, se presenta el siguiente listado de casos donde se aplicó con éxito estas técnicas y metodologías dentro del rubro bancario o similar. Cada uno se desarrolla en el presente proyecto académico bajo la estructura organizada de contexto y problemática, metodología y herramientas, resultados y conclusiones.

#### 3.1. Primer Caso

El primer caso corresponde a la publicación académica titulada *Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering*.

##### 1. Contexto y problemática

El autor explica que los clientes de la banca por Internet se han visto incrementados rápidamente, en especial desde 2012 hasta 2014, donde la cantidad de interacciones con este tipo de plataformas se ha visto incrementado en 1 trillón por año. Asimismo, la minería de datos aplicada al sector bancario se ha orientado a la retención de clientes, la prevención de fraudes y su detección y la segmentación de clientes.

Adicionalmente, menciona que las metodologías *K-means* y *K-medoids* han sido usadas para aplicaciones de *text mining* anteriormente para comparar su desempeño en la clusterización de textos. De estas prácticas, se obtuvo que *K-medoids* superó a *K-means* en el campo de los textos y el estadístico.

Asimismo, el objetivo de la investigación es el de realizar la segmentación de los clientes del banco XYZ usando el método *K-means* y *K-medoids* basado en su puntaje RFM (*Recency, Frequency, Monetary*) sobre las transacciones mediante la banca por Internet y; de esa manera, comparar el desempeño de ambos métodos.

##### 2. Metodología y herramientas

La metodología empleada es la nombrada como *Knowledge Discovery*, la cual consta de cinco etapas: limpieza de datos, selección de datos, transformación de datos, minería de datos y el cálculo del desempeño de la clusterización.

La primera etapa señala los atributos de la información disponible de las transacciones: fecha de transacción, fecha de publicación, número de cuenta, identificador del cliente, código de tarjeta de débito, monto de la transacción y el balance. El problema detectado fue la localización de datos no numéricos en los dos últimos campos de la data. No obstante, esto se solucionó eliminando las partes decimales de toda suma numérica de esos campos.

La segunda etapa radica en la selección de datos. En ella, fueron seleccionados los campos de fecha de transacción, número de cuenta, identificador del cliente y monto de la transacción para el desarrollo del método RFM para la segmentación del cliente.

La tercera etapa se refiere a la transformación de los datos con la aplicación del modelo. La información fue recopilada en torno al identificador de cada cliente para obtener las variables del modelo.

- Recencia: intervalo entre la última fecha de transacción y el último día de diciembre de 2015.
- Frecuencia: cuántas transacciones fueron realizadas en 2015.
- Monetario: el importe total de las transacciones en 2015.

La cuarta etapa fue enfocada en la minería de datos. La primera parte aplica *K-Means (KM) Clustering* con un *dataset* de  $n$  *datapoints* y un número  $k$  como el número de clústeres. El algoritmo de fraccionamiento se encarga de organizar los datos en clústeres independientes, donde  $k \leq n$ . La segunda parte se centra en aplicar *K-Medoids (K-Med) Clustering* que, a diferencia de la metodología anterior, esta no minimiza el error cuadrático, sino que minimiza la suma de las diferencias entre los *datapoints* dentro de un clúster y selecciona un dato por agrupación como su centro (*medoid*).

La quinta etapa se centra en calcular el desempeño de las técnicas de clusterización de dos formas. La primera es el AWC (*Average Within Cluster*), que se calcula con la siguiente expresión donde  $Y_i$  es el valor de cada *datapoint* dentro del clúster y  $\bar{Y}_i$  es el centroide del clúster.

$$SSE = \sum (Y_i - \bar{Y}_i)^2$$

La segunda es el DBI (Davies-Bouldin Index). Este indicador se calcula como sigue donde  $n$  es el número de clústeres,  $\sigma_i$  es el promedio de las distancias entre el centroide y los elementos del clúster  $i$  y  $\sigma_j$  es el promedio de las distancias entre el centroide y los elementos del clúster  $j$ . Por último,  $d(c_i, c_j)$  representa la distancia entre los centroides del clúster  $i$  y el clúster  $j$ .

$$DB = \frac{1}{n} \sum_{i=1, i \neq j}^n \max \left( \frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

El buen desempeño de una clusterización se mide por cuan pequeño es el valor de estos indicadores.

### 3. Resultados

Los resultados de los dos indicadores desarrollados se detallan en la Tabla 10 a continuación. Esta muestra que, según el AWC, el desempeño de KM fue superior a K-Med. En cambio, según el DBI, la superioridad de KM se debió a una ligera diferencia respecto de K-Med, lo que permite constatar que ambos métodos de clusterización poseen una buena performance bajo este indicador. Asimismo, solo observando el DBI, se vislumbra que el número óptimo de clústeres es 2.

Este resultado no estuvo acorde a la investigación. Por ello, se construyó un modelo basado en segmentos RFM donde se hallaron 6 categorías de clientes nombrados como

*Superstars, Golden customers, Typical customers, Exceptional Occasions customers, Everyday shoppers, y Dormant customers.*

Tabla 10 *Clustering Performance*

k	AWC		DBI	
	KM	K-Med	KM	K-Med
2	3.045	7.771	0.935	0.986
3	2.310	6.195	1.052	1.363
4	1.717	4.664	0.976	1.879
5	1.504	2.813	1.006	1.400
6	1.321	2.686	1.071	1.350
7	1.147	2.151	1.001	1.203
8	1.0267	1.839	0.941	1.736
9	0.959	1.687	0.982	1.329
10	0.941	1.527	$-\infty$	1.366

Tomado de “*Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering*”, por Aryumi, Didik y Miranda, 2018.

Este resultado no estuvo acorde a la investigación. Por ello, se construyó un modelo basado en segmentos RFM donde se hallaron 6 categorías de clientes nombrados como *Superstars, Golden customers, Typical customers, Exceptional Occasions customers, Everyday shoppers, y Dormant customers.*

Las características de cada agrupación relacionadas con la aplicación de KM y K-med se observa en la Tabla 11. De esta manera, los grupos encontrados son posibles de tratar como agrupaciones con características de comportamiento similares para campañas de *Marketing.*

#### 4. Conclusiones

La performance de la clusterización fue medida y comparada. El método *K-means* superó al *K-medoids* según el AWC. En contraste, usando el DBI, *K-means* fue solo ligeramente superior a *K-medoids*.

Asimismo, el número óptimo de clústeres óptimo es 2 basado en el menor valor de DBI. Además, se recomienda que para futuros trabajos de este tipo es preciso combinar la información de las transacciones bancarias por Internet, información sociodemográfica e información del producto y; crear una regla de comportamiento para cada segmentación del cliente.

Tabla 11 Características del clúster

RFM segments	KM		K-Med	
	Number of customers/centroid	Range of RFM score	Number of customers/centroid	Range of RFM score
Superstars -Highest F -Highest M	685/ R: 4.524 F: 4.623 M: 4.442	12-15	589/ R: 2 F: 5 M: 4	7-14
Golden customers -High F -Second highest M	374/ R: 2.257 F: 4.377 M: 3.979	7-13	953/ R: 5 F: 4 M: 3	8-15
Typical customers -Average F -Average M	411/ R: 2.071 F: 2.336 M: 3.886	6-11	234/ R: 4 F: 1 M: 3	7-12
"Exceptional Occasions" customers -Second lowest F	448/ R: 3.395 F: 1.844 M: 1.819	5-9	213/ R: 4 F: 1 M: 1	5-9
"Everyday" Shoppers -Increased F - Medium to low M	466/ R: 4.455 F: 3.144 M: 2.758	9-12	30/ R: 5 F: 1 M: 2	8-9
Dormant customers -Low R -Low F -Low M	659/ R: 1.349 F: 1.340 M: 1.366	3-6	1024/ R: 2 F: 1 M: 2	3-8

Tomado de "Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering", por Aryumi, Didik y Miranda, 2018.

### 3.2. Segundo Caso

El segundo caso corresponde al *paper* titulado *Application of Social Media Analytics in the banking sector to drive growth and sustainability: A proposed integrated framework* publicado en el texto compilatorio de la *2018 Open Innovations Conference* desarrollada en Sudáfrica.

#### 1. Contexto y problemática

El Análisis de Redes Sociales (SMA, por sus siglas en inglés) se compone de herramientas relacionadas a un marco de trabajo utilizado por los bancos y compañías de otros rubros para medir y evaluar el desempeño de sus plataformas en redes sociales destinadas a alcanzar objetivos específicos del negocio.

Un enfoque de este campo está interesado en el desarrollo y evaluación de herramientas informáticas y metodologías para monitorear, recolectar, resumir, analizar y visualizar la información extraída de las redes sociales bajo ciertos requerimientos asociados al negocio.

La sola presencia de una web es insuficiente cuando existe una necesidad de entender cómo sacar provecho de la información generada en las redes sociales mediante el análisis de estas. Por ello, es imprescindible definir y adoptar métricas para hacer posible la medición del éxito del banco.

El SMA está definida como el arte y la ciencia de extraer información escondida, pero valiosa, a partir de grandes cantidades de datos estructurados o semiestructurados. De esta manera, la empresa que adopta esta práctica es permitida de tomar decisiones importantes al estar bien informada. Las técnicas comúnmente usadas para aplicar el análisis de redes sociales son la minería de datos, el análisis de textos y el análisis de sentimientos.

Este tipo de análisis les brinda a los bancos la oportunidad de elaborar campañas publicitarias específicas para determinados grupos de clientes. Esto incluye una mejoría en la administración de las relaciones sociales con el consumidor, la identificación de clientes con potencial y el aseguramiento del servicio tan pronto se encuentre con las necesidades del cliente. Además, la aplicación de inteligencia para la retención de clientes puede ser usada para responder a sus quejas generadas a partir de sus encuentros con malas experiencias con los productos y servicios ofrecidos por una entidad financiera.

Como ejemplo de la integración y muestra de la complementariedad de la banca y el mundo digital, se presenta el caso del banco sudafricano ABSA que lanzó una plataforma de

*ChatBanking* vía Twitter en mayo de 2016 y otra en Facebook, unos meses después. Estos ambientes permitieron a los consumidores revisar sus estados de cuenta, realizar transacciones y hacer pagos por servicios de una forma muy simple. El mayor logro de esta iniciativa ha sido el de asegurar la comodidad de sus clientes independientemente de la ubicación en la que se encuentren.

## 2. Metodología y herramientas

Este documento se enfoca en realizar una investigación cualitativa con un enfoque que envuelve una extensiva revisión de literatura relacionada al SMA como artículos de revistas con una temática acorde, *papers*, libros de texto y actas de congresos.

El objetivo del *paper* es proponer un marco de trabajo integrado destinado a ayudar a las instituciones financieras a generar valor a partir del análisis de redes sociales para, de este modo, potenciar el crecimiento del negocio y su sostenibilidad.

Los ocho temas recogidos por el autor son *topic modelling*, *opinion mining*, *sentiment analysis*, *social network analysis*, *trend analysis*, *popularity prediction*, *customer engagement* y *visual analytics*. Cada uno tiene un rol crítico en las operaciones de las entidades bancarias.

## 3. Resultados

El autor inicia con una recopilación sobre la teoría de cada herramienta de SMA, identificando su área de influencia, su definición y su propósito.

- *Topic modelling*. Es usado en política, salud pública y negocios para detectar temas dominantes recurrentes en medio de una gran cantidad de texto con el propósito de usar estadística avanzada y aprendizaje de máquina para identificar los tópicos latentes en tiempo real.
- *Opinion mining*. Presente en el *E-commerce*, las finanzas, los recursos humanos y la mercadotecnia. Se diferencia del *Sentiment analysis* al enfocarse preferentemente en los puntos de vista, las creencias y el juicio crítico por sobre los sentimientos. De esta

manera, mide estos parámetros en base al criterio que se determine por el enfoque del análisis.

- *Sentiment analysis*. Se usa en el *E-commerce*, política, mercadotecnia, servicios, cadena de suministros y finanzas. Al contrario de la *Opinion Mining*, esta herramienta se orienta a una interpretación más profunda de los comentarios de los clientes y consumidores sobre sus sentimientos, evaluaciones, apreciaciones, actitudes y emociones sobre los productos, servicios, organizaciones, colaboradores, problemas, eventos, temas y atributos del negocio. Tiene como fin la medida de las emociones individuales, grupales y comunitarias hacia cualquier tipo de evento relacionado a la empresa.
- *Social Network Analysis*. Ampliamente utilizado en la CRM, la política, la mercadotecnia, en áreas funcionales como Producción y Operaciones. Su funcionamiento depende de la conexión que tenga un determinado individuo, a modo de nodo, con otros nodos de similar interés, conocimiento u opinión. Su propósito es medir los tipos y la profundidad de la relación de los nodos dentro de las redes.
- *Trend analysis*. Aplicable al estudio del consumidor, las ventas y la efectividad de una campaña publicitaria. Básicamente, permite realizar el pronóstico de la tendencia del mercado o el comportamiento del consumidor a partir de información histórica. El último fin de esta herramienta es elaborar proyecciones de ventas, participación de mercado, crecimiento de clientes y movimientos en la bolsa de valores basadas en series de tiempo y regresión.
- *Popularity prediction*. Tiene presencia en el área de los negocios, la mercadotecnia, las relaciones públicas, el entretenimiento y la política. Por su definición, es un método de recolección de retroalimentación positiva y negativa sobre determinados personajes o eventos para analizar su nivel de popularidad y predecir su comportamiento basado en

la evidencia. Este ejercicio permite a la organización pronosticar la demanda de un producto, servicio o evento.

- *Customer engagement analysis*. Resulta útil para el E-commerce, la mercadotecnia, los negocios, la política y el entretenimiento. Se define como el proceso para prolongar la interacción de los usuarios de una red social alrededor de un tema de conversación o evento mediante incentivos apropiados y el entendimiento de las características del consumidor en Internet. El objetivo es medir el éxito de las actividades *online* como una campaña comercial sin fines de lucro para, de esa manera, ayudar a la organización a entender la coyuntura y tomar acción para ganar éxito en el entorno virtual.
- *Visual analytics*. Se presenta en la mercadotecnia, las ventas, el *E-commerce* y el *Big data*. Corresponde a un proceso iterativo que abarca la recolección de información, su procesamiento y una posterior toma de decisiones. El propósito es usar interfaces gráficas para presentar, explorar y corroborar relaciones entre variables.

Una vez acotadas estas herramientas del SMA, el autor presenta su propuesta de marco de trabajo integrado de análisis de redes sociales como se muestra en la **¡Error! No se encuentra el origen de la referencia.** En la ilustración, se listan los ocho métodos de análisis, los cuales se interrelacionan en una segunda etapa donde se enumeran los prerequisites para el análisis final que conciernen al estudio de las fuentes de valor, construir un ecosistema de datos, el modelamiento de ideas, su transformación en acciones y la implementación de una medida para producir resultados. Por último, las estrategias adoptadas por la organización desembocan en la mejoría del crecimiento del negocio y su sostenibilidad.

#### 4. Conclusiones

El autor concluye con una reflexión sobre el marco de trabajo propuesto, remarca la importancia de este sistema sobre las instituciones bancarias que han implementado tecnologías con el propósito de realizar un análisis de redes sociales, el cual pueden usar para generar valor

para sus negocios. Asimismo, acota que esta metodología no recomienda alguna plataforma de red social en específico, ya que esa decisión será distinta en concordancia con las características del banco, su posicionamiento geográfico y su público objetivo.

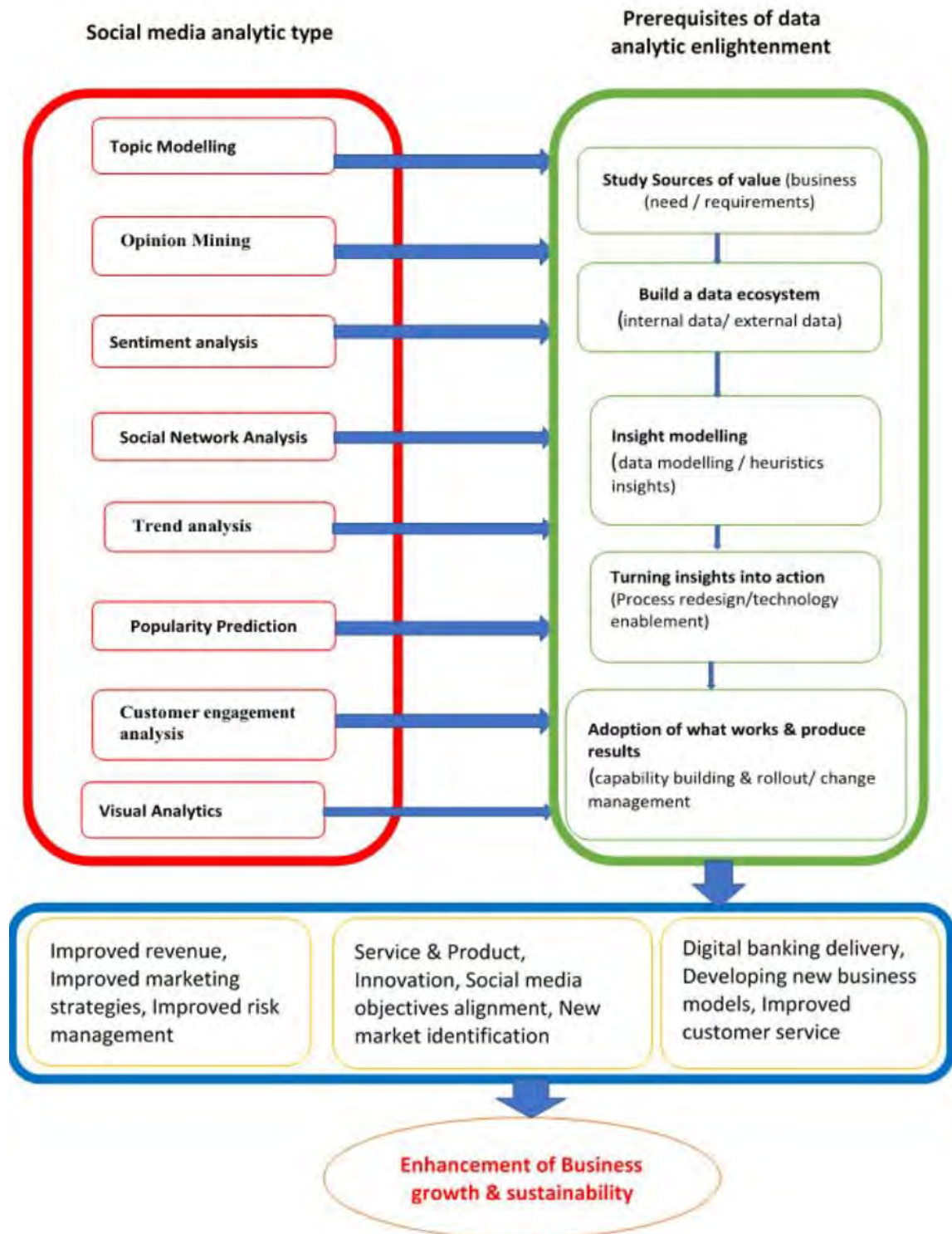


Figura 13 Marco de trabajo integrado propuesto sobre el SMA

Tomado de “*Application of Social Media Analytics in the banking sector to drive growth and sustainability: A proposed integrated framework*”, por Bankolem y Manzira, 2018.

En las siguientes líneas, acota que los bancos están constantemente realizando esfuerzos para explorar otras formas de alcanzar el crecimiento de las ganancias y la eficiencia operacional, campos donde la analítica contribuye ampliamente si es identificada y aplicada racionalmente. Además, menciona que, a pesar de que el marco propuesto está orientado a ser genérico, se podría enfocar este a ser reorientado para distintos tipos de banco como el de moneda extranjera, los industriales, los de agricultura, los comerciales, los de inversión y los centrales.

### 3.3. Tercer Caso

El tercer caso trata del *paper* titulado *Calculating the required cash in bank branches: a Bayesian-data mining approach*. El texto fue publicado en 2017 y, posteriormente, recopilado en la revista *Neural Computing & Applications* en 2018.

#### 1. Contexto y problemática

En el mundo de la banca, una de las preocupaciones más significativas de un gerente, particularmente la de un gerente de sucursal, al momento de tomar decisiones es el aseguramiento de una cantidad suficiente de dinero en efectivo en las agencias. Si una agencia no cuenta con el monto necesario, esta no es capaz de responder a las necesidades de los clientes, por lo tanto, no quedarán satisfechos con el servicio ofrecido. Por otro lado, la existencia de un monto superior al requerido en una filial genera que, al final del día, el remanente sea transferido a la tesorería, lo cual ocasiona que los costos por la depreciación de efectivo y el uso de vehículos de transporte de dinero se incrementen. Asimismo, esta última situación suscita que este recurso no sea capaz de ser usado para generar ganancias. Por esta razón, los bancos intentan optimizar el flujo de dinero de sus sucursales basándose en su ejercicio diario. Ante esta coyuntura, la solución se plantea como un método apropiado y eficiente para estimar el dinero en efectivo requerido por cada dependencia dentro de un periodo de tiempo para prevenir estos inconvenientes.

## 2. Metodología y herramientas

La minería de datos se perfila como la metodología más reciente y más aplicada para este propósito, tal y como describe el autor en su obra, pues hace uso de métodos estadísticos no paramétricos. Además, es útil al enfrentarse a varios datos y porque la información recolectada no acostumbra a satisfacer la mayoría de las antiguas hipótesis estadísticas como la de poseer una distribución específica. Al no observarse este tipo de situaciones, se originan problemas debido a la autenticidad y la precisión de los resultados de la investigación. Dos de las más importantes y confiables técnicas de este tipo son el *clustering* y las redes neuronales, justamente las empleadas en el documento.

En esta línea, el análisis bayesiano multivariado es considerado una técnica importante en la estadística. Usando su enfoque en metodologías como las dos mencionadas resulta en modelos y cálculos de mayor precisión y efectividad. Asimismo, es idóneo para lidiar con el sobreajuste y optimizar la regularización en el proceso de entrenamiento de las redes.

A través de la utilización del enfoque bayesiano, se estima el dinero en efectivo requerido por las agencias de un banco iraní por medio de la proyección del ingreso y salida de dinero de las dependencias. Para este propósito, se considera el hecho de que cada filial es distinta y todas ellas pueden ser agrupadas usando dos métodos de *clustering*, uno al estilo clásico y el otro a la manera bayesiana, reuniendo a las sucursales con características similares. A continuación, se aplica la metodología de redes neuronales, una con el método clásico y otra con el enfoque bayesiano. Por último, considerando los resultados obtenidos en ambos procesos, se puede determinar el mejor modelo para calcular la dotación de efectivo de la oficina.

La población estadística de este *paper* incluye las sucursales de un banco en Tehran, la capital de Irán, que, debido a la variedad de puntos en todas sus regiones, se ha aplicado el método de muestreo aleatorio para seleccionar la muestra de agencias.

La información como los días trabajados, los días a la semana, el salario diario, el subsidio, las ganancias, los feriados y eventos oficiales han sido extraída de 18 agencias en el periodo comprendido entre el 21 de abril de 2014 y el 22 de septiembre del mismo año.

La peculiaridad del *clustering* bayesiano es que requiere de cálculos avanzados de simulación mediante el método de *Markov Chain Monte Carlo*. Esta premisa junto a la estimación del ingreso y salida de efectivo son conducidas por la aplicación de redes neuronales bayesianas usando el software R.

### 3. Resultados

La primera parte corresponde a los resultados del *clustering* jerárquico de las agencias. Por un lado, se usó el coeficiente de similitud de Gower y el algoritmo del vecino más lejano para crear las agrupaciones. Estas fueron validadas por el índice de Rousseeuw's Silhouette, dando como resultado que 11 es la cantidad óptima de *cluster* como se aprecia en la Tabla 12 donde su índice es el de mayor valor.

Tabla 12 Valores del índice de evaluación de validez de las distintas cantidades de *cluster*

Numbers of clusters ( <i>K</i> )	2	3	4	5	6	7	8
Rousseeuw's Silhouette	0.361	0.351	0.364	0.365	0.340	0.340	0.365
Numbers of clusters ( <i>K</i> )	9	10	11	12	13	14	15
Rousseeuw's Silhouette	0.386	0.411	0.412	0.408	0.371	0.289	0.262

Tomado de “*Calculating the required cash in bank branches: a Bayesian-data mining approach*”, por Baghbani y Eskandari, 2018.

Por otro lado, los resultados al aplicar el enfoque bayesiano a la conformación de los *cluster* arrojan que son 3 las agrupaciones óptimas al haber alcanzado esta cantidad el valor más alto en la distribución de logaritmo de probabilidad posterior como se muestra en la Figura 14.

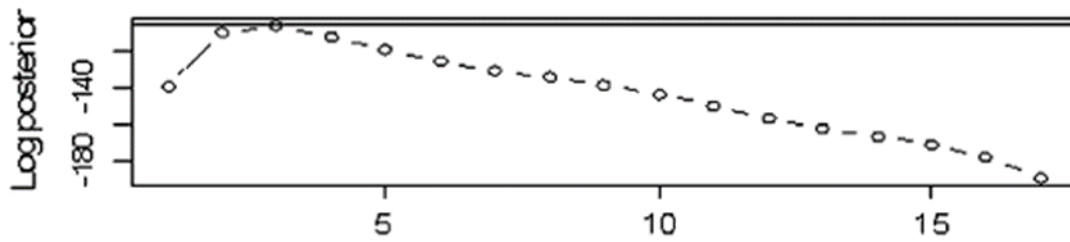


Figura 14 Valores de la distribución del logaritmo posterior en base a la cantidad de *cluster*

Tomado de “*Calculating the required cash in bank branches: a Bayesian-data mining approach*”, por Baghbani y Eskandari, 2018.

La Tabla 13 resume la asignación de cada filial a un determinado *cluster* de acuerdo a los dos métodos utilizados. El diferenciados de cada observación es su código de agencia.

Tabla 13 Cantidad de *cluster* de las sucursales basada en los métodos de *clustering*

Branch code	Cluster number based on hierarchical clustering	Cluster number based on Bayesian clustering
450	1	1
66	2	3
107	2	3
135	3	3
165	4	2
54	4	2
317	5	1
411	6	2
21	7	2
312	6	2
362	2	3
392	3	3
706	2	3
1019	8	2
0	7	2
56	9	1
231	10	2
270	11	1

Tomado de “*Calculating the required cash in bank branches: a Bayesian-data mining approach*”, por Baghbani y Eskandari, 2018.

En la siguiente sección, se calculan los resultados de la red neuronal bayesiana a partir de los datos originados por el *clustering*. En primera instancia, para estimar el dinero que ha de entrar a la agencia usando los resultados del agrupamiento bayesiano, se consideran primero dos neuronas en la capa oculta. Luego es contrastada la cantidad de neuronas con el error

cuadrático medio (RMSE, por sus siglas en inglés), con lo cual, se seleccionan finalmente 18 neuronas en la capa oculta, cuya estructura de red posee el menor error. El mismo ejercicio fue realizado para el caso del dinero que sale de la agencia y resultó óptima la cantidad de 14 neuronas. En segunda instancia, se continúa usando la red neuronal bayesiana, pero ahora aplicada a los resultados del *clustering* clásico. El resultado para la estimación del dinero que ingresa a la dependencia arrojó 15 neuronas en la capa oculta. Mientras que, para el dinero que sale de la oficina, se obtuvieron 7 neuronas con el menor RMSE.

El dinero en efectivo requerido en las sucursales se obtiene de la diferencia entre el flujo de entrada y el de salida estimado por las redes neuronales. Por ello, la Figura 15 muestra esta diferencia real de dinero en 18 agencias durante 5 meses comparada con los resultados de la estimación elaborada por el método bayesiano de redes neuronales aplicado al *clustering* clásico y el bayesiano. Es notable que el resultado alcanzado por el *clustering* bayesiano es más cercano al flujo real que por el método regular.

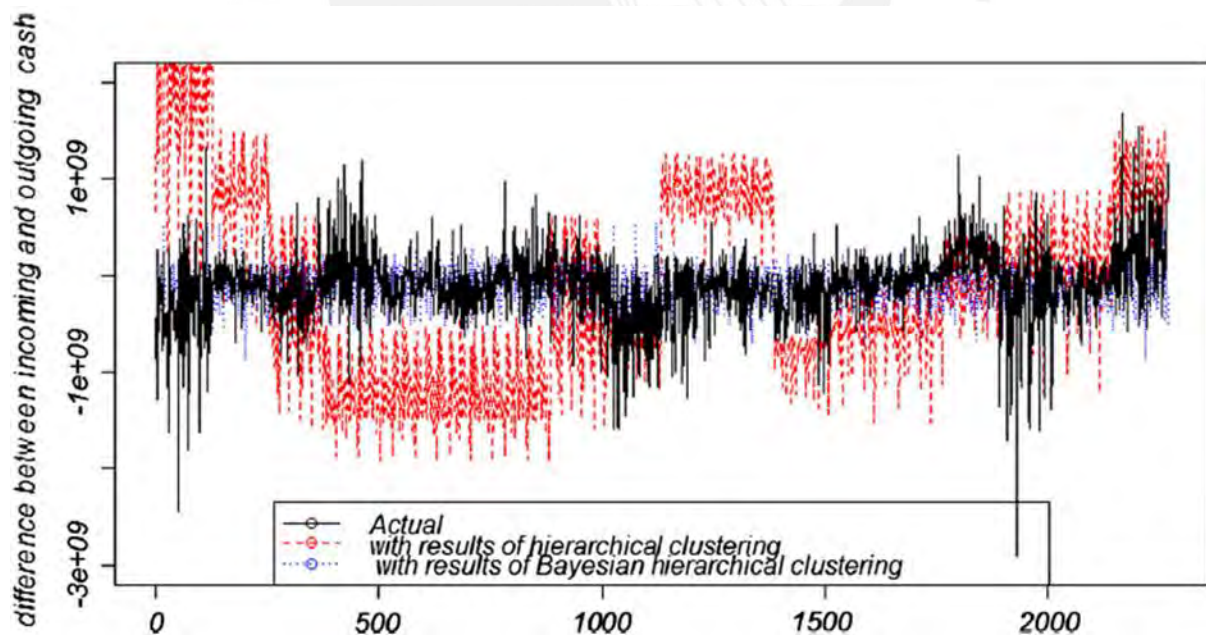


Figura 15 Diferencia entre el flujo real de entrada y salida de dinero de 18 agencias durante 5 meses comparada con la estimación por el método bayesiano de redes neuronales considerando los resultados obtenidos del *clustering* clásico y el bayesiano

Tomado de “*Calculating the required cash in bank branches: a Bayesian-data mining approach*”, por Baghbani y Eskandari, 2018.

La sección restante corresponde a la aplicación clásica de redes neuronales sobre ambos tipos de agrupación. El 70 % de la información forma parte de la red de entrenamiento y el 30 % restante es para el testeo. A continuación, esta información es evaluada de forma similar al apartado anterior con la particularidad de que los datos fueron entrenados con 500 iteraciones y un algoritmo de propagación hacia atrás con un ratio de aprendizaje de 0,2. Por lo tanto, la cantidad de neuronas en la capa oculta determinada para el *clustering* jerárquico y para el bayesiano, sobre el monto de entrada y el de salida, respectivamente, es de 20, 19, 19 y 17.

De forma similar, se presenta la Figura 16 donde se indica el mismo espectro de la ilustración anterior, pero aplicada bajo el enfoque clásico de las redes neuronales. Asimismo, resalta el hecho de que los resultados del *clustering* bayesiano sean más cercanos al comportamiento de los datos reales del rango evaluado.

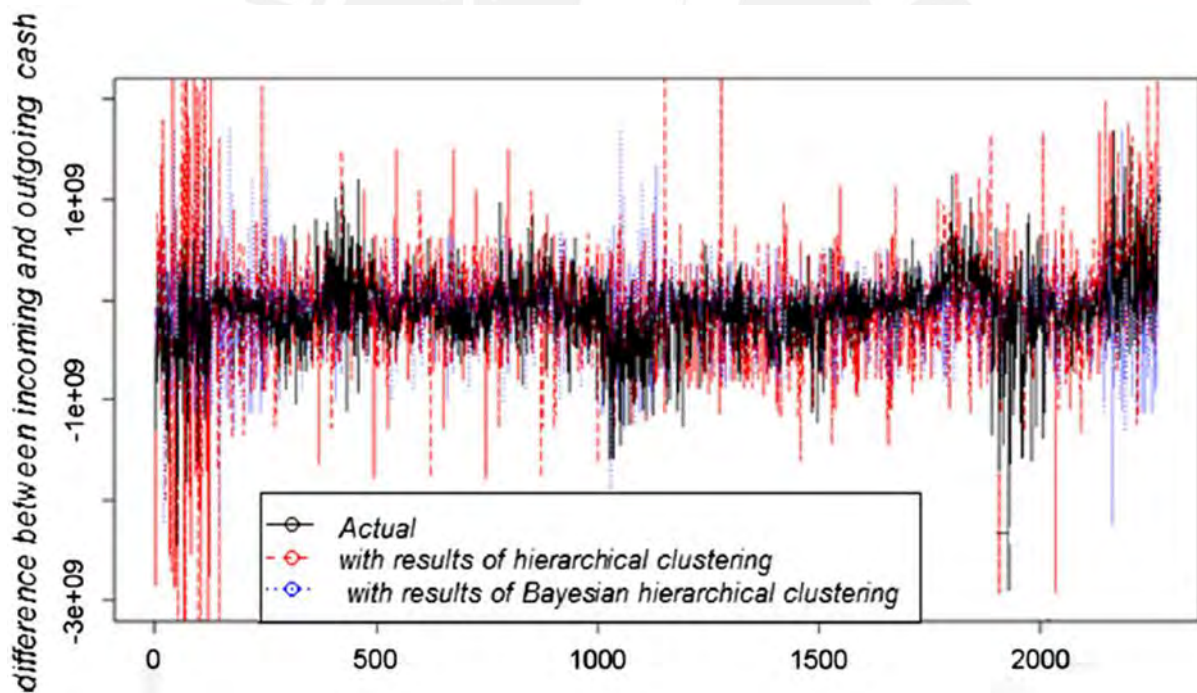


Figura 16 Diferencia entre el flujo real de entrada y salida de dinero de 18 agencias durante 5 meses comparada con la estimación por el método clásico de redes neuronales considerando los resultados obtenidos del *clustering* clásico y el bayesiano

Tomado de “Calculating the required cash in bank branches: a Bayesian-data mining approach”, por Baghbani y Eskandari, 2018.

De las últimas dos ilustraciones se recoge que el desempeño de la red neuronal bayesiana considerando los resultados del *clustering* bayesiano es superior al del resto de combinaciones y es el mejor modelo para obtener el dinero en efectivo requerido en cada sucursal del banco iraní. Esta afirmación es respaldada por la Tabla 14 que cuantifica la desviación media absoluta (MAD, por sus siglas en inglés) y el error cuadrático para la diferencia del ingreso y la salida de dinero con las 4 combinaciones.

Tabla 14 Criterios de error para los resultados de los cálculos

Mean absolute deviation (MAD)	Root-mean-square error (RMSE)	Values in Million Rials
348	484	Error value in difference between input and output cash of branches estimated through classic neural network based on Bayesian clustering results
447	681	Error value in difference between input and output cash of branches estimated through classic neural network based on classic clustering results
287	393	Error value in difference between input and output cash of branches estimated through Bayesian neural network based on Bayesian clustering results
859	1134	Error value in difference between input and output cash of branches estimated through Bayesian neural network based on classic clustering results

Tomado de “*Calculating the required cash in bank branches: a Bayesian-data mining approach*”, por Baghbani y Eskandari, 2018.

#### 4. Conclusiones

De las 4 combinaciones generadas a partir del estudio de los métodos clásico y bayesiano de redes neuronales y sus similares en torno al *clustering*, el correspondiente a la aplicación del enfoque bayesiano en ambas herramientas presenta una mejor performance al evaluarse el MAD y el RMSE, obteniendo los menores valores en cada uno, equivalentes a 287 millones y 393 millones de riales iraníes.

#### 4. Resumen Metodológico

Esta sección se encarga de resumir la metodología empleada por cada capítulo de la tesis para su desarrollo y obtención válida de resultados.

##### 1. Capítulo 2. Descripción y diagnóstico del sistema

Se realiza la descripción de la organización bajo análisis tocando los puntos más importantes como la actividad comercial que desarrollan, los productos o servicios que ofrece, los clientes, los proveedores, el estudio organizacional y el análisis coyuntural, dando hincapié en la situación de la empresa bancaria frente al crecimiento de la competencia durante las últimas temporadas. El diagnóstico de la organización se focaliza en encontrar, en primera instancia, la causa raíz de los problemas sobre los indicadores de gestión de las actividades del subproceso de mayor importancia que parte de la evaluación previa de los macroprocesos y procesos de mayor prioridad para la entidad financiera. Se utilizan las herramientas indicadas sobre las herramientas de diagnóstico con el objetivo de determinar la contramedida idónea para el desarrollo de la propuesta de mejora.

##### 2. Capítulo 3. Propuesta de mejora

Se desarrolla la propuesta de mejora basado en la contramedida identificada como mejor durante el capítulo anterior. El análisis exploratorio se realiza para comprender en mejor medida el proceso en general donde se ha de implementar el modelo. Es en este capítulo donde sucede la materialización del modelo programa en lenguaje R. La metodología busca establecer las bases del proceso en un inicio, para luego presentar los datos, validarlos, tomar una muestra representativa y proceder con el modelo orientado a los objetivos determinados por la tesis.

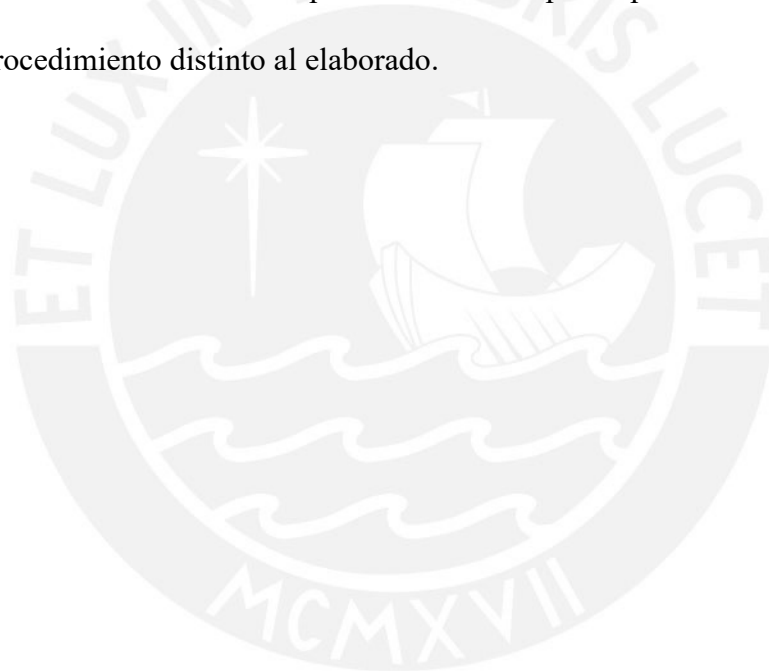
##### 3. Capítulo 4. Evaluación técnica y económica

Se evalúa la viabilidad de esta propuesta a nivel económico, técnico y de gestión con la finalidad de justificar el riesgo a tomar por el modelo y demostrar el impacto positivo que desencadenará en el banco desde el retorno sobre la inversión (ROI) hasta la satisfacción del

cliente, cuyo detalle se presenta con un grupo de indicadores orientados al control de las agencias bajo el nuevo foco de distribución. Asimismo, se evalúan los indicadores de rentabilidad VAN y TIR con el fin de cuantificar los ahorros e ingresos estimados a partir de la propuesta.

#### 4. Capítulo 5. Conclusiones y recomendaciones

Con base a los hallazgos y resultados obtenidos a lo largo del trabajo académico, se procede a la presentación de las conclusiones y recomendaciones evocadas por cada capítulo. La intención de las recomendaciones es la de buscar un mejor grado de satisfacción para estudios posteriores relacionados con la presente donde se pueda plantear una mejora con una herramienta o procedimiento distinto al elaborado.



## Conclusiones

Con base en el desarrollo del presente documento, es posible concluir lo dictado en los enunciados de la siguiente lista.

- El trabajo de investigación sirve como el soporte para la realización del trabajo de tesis homónimo, donde se aplican las herramientas detalladas en el informe. Las primeras están orientadas a elaborar el diagnóstico de la entidad bancaria tras analizar la estructura organizacional de la empresa y obtener la causa raíz que genere una contramedida adecuada donde proponer una mejora. Por otro lado, las segundas corresponden a la manera de abordar el problema otorgando la metodología para presentar y probar una propuesta de mejora adecuada que, en este caso, se encuentra orientada a la aplicación de *Data Science*.
- De acuerdo con el caso consultado, la performance del método de *clustering K-means* supera al *K-medoids* según los indicadores AWC y DBI, de manera ligera. Esto supone un contraste eventual durante el desarrollo de la tesis al probar las propuestas de mejora.
- De acuerdo con el segundo caso consultado, el análisis de redes sociales es capaz de generar valor para los negocios que lo apliquen (banca, industria, agricultura, comercio, inversión, etc.). Asimismo, se acota que esta metodología no recomienda alguna plataforma de red social en específico, ya que esa decisión será distinta en concordancia con las características del banco, su posicionamiento geográfico y su público objetivo.
- De acuerdo con el tercer caso, la aplicación de redes neuronales en el entorno de las finanzas es posible y reduce el trabajo realizado por la aplicación combinada del enfoque bayesiano para el desarrollo de la red neuronal y el *clustering* previo. Esto se fundamenta en que ambas herramientas presentan una mejor performance al evaluarse el MAD y el RMSE, obteniendo los menores valores en cada uno, equivalentes a 287 millones y 393 millones de riales iraníes.

## Bibliografía

- Aiteco Consultores. (s.f.). *Mapa de Procesos. Gestionar los Procesos*. Obtenido de Aiteco Consultores Desarrollo y Gestión: <https://www.aiteco.com/origen-del-mapa-de-procesos/>
- Arcturus7. (2019). *Quality Function Deployment – QFD2000 Pro*. Recuperado el 22 de noviembre de 2019, de Arcturus 7 Shop: <https://www.arcturus7.com/a7s/product/quality-function-deployment-qfd2000/>
- Aryumi, M., Didik, E., & Miranda, E. (2018). Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering. *2018 International Conference on Information Management and Technology (ICIMTech)*, 412-416. Recuperado el 25 de noviembre de 2019, de <https://ieeexplore-ieee.org.ezproxybib.pucp.edu.pe/stamp/stamp.jsp?tp=&arnumber=8528086>
- Baghbani, G., & Eskandari, F. (2018). Calculating the required cash in bank branches: a Bayesian-data mining approach. *Neural Computing and Applications*, 30(9), 2831-2841. Recuperado el 17 de abril de 2020, de <https://doi-org.ezproxybib.pucp.edu.pe/10.1007/s00521-017-2888-9>
- Bankolem, F., & Manzira, F. (2018). Application of Social Media Analytics in the banking sector to drive growth and sustainability: A proposed integrated framework. *2018 Open Innovations Conference*, 223-233. Recuperado el 14 de abril de 2020, de <https://ieeexplore-ieee.org.ezproxybib.pucp.edu.pe/stamp/stamp.jsp?tp=&arnumber=8535833>
- Bernal, L., Dornberger, U., & Suvelza, J. A. (2009). *Quality Function Deployment (QFD) para servicio*. Recuperado el 22 de noviembre de 2019, de CONOSCOPE: [https://www.conoscope.de/fileadmin/user\\_upload/Downloads/Guia\\_QFD\\_Servicios.pdf](https://www.conoscope.de/fileadmin/user_upload/Downloads/Guia_QFD_Servicios.pdf)

Betancourt Quintero, D. F. (24 de noviembre de 2018). *Cómo hacer una matriz de priorización*.

Recuperado el 22 de noviembre de 2019, de Ingenio Empresa:  
<https://ingenioempresa.com/matriz-de-priorizacion/>

Cadena Echeverría, J. L. (2016). Guía para el diseño y documentación de procesos. *Yura:*

*Relaciones internacionales*(6), 57-83. Obtenido de  
[http://world\\_business.espe.edu.ec/wp-content/uploads/2016/03/Art%C3%ADculo-6.4-Gu%C3%ADa-de-procesos.pdf](http://world_business.espe.edu.ec/wp-content/uploads/2016/03/Art%C3%ADculo-6.4-Gu%C3%ADa-de-procesos.pdf)

Escuela Técnica Superior de Ingenieros Industriales. (2006). *El diagrama de Causa-Efecto*.

Vigo: Universidad de Vigo. Recuperado el 22 de noviembre de 2019, de  
<http://gio.uvigo.es/asignaturas/gestioncalidad/GCal0405.DiagramaCausaEfecto.pdf>

Escuela Técnica Superior de Ingenieros Industriales. (2006). *El diagrama de Pareto*. Vigo:

Universidad de Vigo. Recuperado el 22 de noviembre de 2019, de  
<http://gio.uvigo.es/asignaturas/gestioncalidad/GCal0405.DiagramaPareto.pdf>

James, G. (2013). *An Introduction to Statistical Learning with Applications in R*. Nueva York: Springer.

Leebov, W. (1993). *Manual de los administradores de salud para el mejoramiento continuo*.

Manual, Centro de Gestión Hospitalaria, Bogotá.

Madhavan, S. (2015). *Mastering Python for Data Science*. Packt Publishing.

Metodologías Más Soluciones. (2011). *DEI - C - Ficha de Indicador - 2012 11 25*.

Diapositivas, Metodologías Más Soluciones, Lima. Recuperado el 22 de noviembre de 2019, de <https://www.slideshare.net/metodologias/dei-c-ficha-de-indicador-presentacin-2010-10-10>

Milián, E. (s.f.). *Diagrama de Ishikawa*. Universidad de San Carlos de Guatemala. Recuperado

el 23 de noviembre de 2019, de  
<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&>

uact=8&ved=2ahUKEwjpg6Xvn4\_mAhVkoFkKHeoUCz0QFjABegQIAhAC&url=ht  
 tps%3A%2F%2Fenosiyivol.jimdo.com%2Fapp%2Fdownload%2F11313774895%2F  
 5.%2BDIAGRAMA%2Bde%2Bishikawa.pdf%3Ft%3D1558068435&usg

Ministerio de Planificación Nacional y Política Económica. (2009). *Guía para la Elaboración de Diagramas de Flujo*. Manual, Ministerio de Planificación Nacional y Política Económica, Área de modernización del Estado, San José. Recuperado el 22 de noviembre de 2019, de <http://biblioteca.udgvirtual.udg.mx/jspui/bitstream/123456789/1223/1/Gu%C3%ADa%20para%20la%20elaboraci%C3%B3n%20de%20diagramas%20de%20flujo.pdf>

Puga Muñoz, M. (2018). *Los cinco por qué (Five Whys)*. Material del curso de Información para las decisiones, Universidad Arturo Prat, Facultad de Ciencias Empresariales, Iquique. Recuperado el 23 de noviembre de 2019, de <http://www.mpuga.com/docencia/Informacion%20Para%20Las%20Decisiones/Los%20Cinco%20por%20ques.pdf>

SAS. (s.f.). *Marketing Analytics*. Recuperado el 24 de noviembre de 2019, de SAS: [https://www.sas.com/en\\_us/insights/marketing/marketing-analytics.html](https://www.sas.com/en_us/insights/marketing/marketing-analytics.html)

Sotelo Seguil, M. G. (2017). *Diagnóstico y mejora para el servicio de la seguridad ciudadana en el distrito de San Martín de Porres*. Tesis para optar el título de ingeniero industrial, Pontificia Universidad Católica del Perú, Lima. Obtenido de <http://hdl.handle.net/20.500.12404/9055>

The R Foundation. (s.f.). *What is R?* Recuperado el 25 de noviembre de 2019, de R: <https://www.r-project.org/about.html>

Toskano Hurtado, G. B. (2005). *El Proceso de análisis jerárquico (AHP) como herramienta para la toma de decisiones en la selección de proveedores: aplicación en la selección del proveedor para la Empresa Gráfica Comercial MyE S.R.L.* Monografía,

Universidad Nacional Mayor de San Marcos, Lima. Obtenido de  
[https://sisbib.unmsm.edu.pe/bibvirtual/monografias/Basic/toskano\\_hg/toskano\\_hg.htm](https://sisbib.unmsm.edu.pe/bibvirtual/monografias/Basic/toskano_hg/toskano_hg.htm)

