

**“PROPUESTA DE INTELIGENCIA DE NEGOCIOS PARA OPTIMIZAR LOS
TIEMPOS EN LA DEPURACIÓN DE LOS DATOS DE CONTACTO Y HÁBEAS DATA
EN LA CAJA DE COMPENSACIÓN FAMILIAR”**

AUTORES:

EDGAR CHAVES RENGIFO - CÓDIGO 1812010327

DIEGO ARMANDO GARCÍA SANABRIA - CÓDIGO 1812010430

SOL ANYI MAYERLY LIMA PAIBA - CÓDIGO 1812010283

EDWIN ADRIAN NARVÁEZ BURBANO - CÓDIGO 1812010279

MERY BIBIANA ROA YOPASA - CÓDIGO 1341980044

ASESOR: MSC. GIOVANNY ALEXANDER BAQUERO VILLAMIL

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO

FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ESPECIALIZACIÓN EN GERENCIA DE PROYECTOS EN INTELIGENCIA DE

NEGOCIOS

BOGOTÁ, D.C. 2018

TABLA DE CONTENIDO

2	Título	3
3	Resumen	3
3.1	Español	3
3.2	Inglés	4
4	Tema	4
4.1	Dedicación	5
5	Fundamentación del proyecto	6
5.1	Marco contextual	6
6	Problema	7
6.1	Diagrama espina de pescado (causa – efecto)	7
7	Justificación	8
7.1	Objetivo General	8
8	Marco conceptual	9
9	Estado del arte	10
10	Objetivos Específicos, actividades y cronograma	18
10.1	Metodología	22
10.2	Presupuesto General del Proyecto	22
11	Viabilidad financiera	23
12	Plan de actividades	23
13	Plan de adquisiciones	24
13.1	Necesidad de adquisiciones externas	24
13.2	Identificación de adquisiciones	24
13.3	Tipos de contrato a utilizar	25
14	Plan de riesgos	25
14.1	Valoración Cualitativa y cuantitativa de riesgos	26
14.2	Monitoreo y seguimiento	27
14.3	Matriz de riesgos aplicada al proyecto	28
15	Plan de interesados	29
15.1	Registro de interesados	29
15.2	Gestión de interesados	30
16	Conclusiones y recomendaciones	30
17	Bibliografía	31

2 Título

Propuesta de inteligencia de negocios, para optimizar los tiempos en la depuración de los datos de contacto y Hábeas Data, en la caja de compensación familiar.

3 Resumen

3.1 Español

Desde el departamento de mercadeo, se realizan diferentes envíos de información comercial a los clientes, ya sea por medio de correo físico, electrónico y mensajes de texto, generando más oportunidad de ventas, con lo cual este proyecto, busca apoyar de manera más rápida y efectiva dicha gestión.

Las bases de datos, que se toman para originar los diferentes envíos de información comercial a los clientes, se reciben en el departamento de mercado de diversas fuentes y en diferentes formatos, las cuales, por lo general vienen incompletas, desordenadas, volviéndose difícil generar una sola base para enviar la información; esto se presenta, porque la consolidación de los datos se realiza en forma manual, haciendo que la entrega de los requerimientos para las diferentes campañas se demore.

Con el fin de agilizar la optimización de la información, se generará en esta propuesta, algunos estándares de información para depurar más fácilmente las direcciones, los correos electrónicos y teléfonos de contacto de las personas que autoricen la utilización de sus datos personales, para el envío de información según la ley 1581 de 2012 “Ley de Hábeas Data”.

Se sistematizará el proceso, creando una serie de macros que ayuden a depurar la información, reduciendo considerablemente los tiempos en los procesos, pasando de un tiempo de entrega de tres y dos semanas, a una semana o menos.

3.2 Inglés

From the marketing department, different shipment of commercial information are made to customers either through physical mail, electronic mail and text messages, generating greater sales opportunities, with which this project seeks to support the management quickly and effectively.

The databases that are taken to originate the different shipments of commercial information to customers are received in the market department from different sources and formats, which are usually incomplete, disorganized, becoming cumbersome and wasteful to create a base to send the information, as result of the manually data consolidation, making the requirement delivery for the different campaigns are delayed.

In order to speed up the information optimization will be created in this proposal, some standards to easily debug the addresses, emails and people telephone numbers who authorize the use of their personal data for the sending information according to the law 1581 of 2012 "Law of habeas data".

The process will be systematized by creating a series of macros that will help to debug the information, reducing the processes time, passing from two or three weeks to one week or less of delivery time.

4 Tema

En el departamento de mercadeo de la caja de compensación, se envían campañas a los clientes, por lo que, para lograr el éxito de las mismas se requiere que la información de la base de datos esté actualizada, sea veraz y confiable adicionalmente la información debe procesarse en el menor tiempo posible con la finalidad de enviar la información de los productos y servicio a los clientes potenciales, la presente propuesta permitirá dar solución a la demora en la entrega de las bases de datos para lograr el contacto con los prospectos o posibles clientes. Lo anterior se logrará realizando unos estándares y creando una serie de macros y análisis de información, que permitan que el proceso de depuración de las bases de datos se realice en menor tiempo, pasando de una depuración manual, a una sistematizada, la cual, ayudará al departamento a agilizar los procesos y

efectivizar la obtención de clientes nuevos. Otro de los beneficios será, que la caja de compensación se evite demandas, por envío de información a personas que han cambiado el estado de autorización de manejo de datos y no desean ser contactadas. Adicionalmente, al tener información actualizada y depurada, será más efectivo el contacto, por lo tanto, se podrán aumentar las ventas de los diferentes bienes y servicios que ofrece la caja dirigiendo los esfuerzos en clientes que mediante el modelo de segmentación se perfilan como prospectos.

Los principales entregables de la depuración de las bases de datos serían:

1. Documento de los estándares para el manejo de las direcciones según la nomenclatura DANE.
2. Macros de depuración de la información sobre correos electrónicos.
3. Macros de depuración de teléfonos.
4. Macro de depuración de direcciones.
5. Plantillas y macros en Excel y Access para la depuración y consolidación de información.
6. Generación de formatos para el envío de correos electrónicos y mensajes de texto.
7. Creación de un modelo que permita segmentar mejor los clientes con el fin de mejorar el envío de información comercial por medio de un perfilamiento adecuado del cliente objetivo para direccionar las estrategias de mercado.
8. Entrega final y cierre del proyecto.

4.1 Dedicación

La siguiente tabla muestra el resumen de la dedicación, que se planea para las principales actividades a desarrollar:

Tipo de Actividad	Sub-actividad	% de Dedicación
Investigación teórica	N/A	20%
Diseño del Proyecto	N/A	25%
Desarrollo	Prototipo/Piloto	40%
	Ambiente de producción	15%

5 Fundamentación del proyecto

5.1 Marco contextual

La caja de compensación es una entidad de Seguridad Social que facilita a sus afiliados, a través de los planes de bienestar, un amplio portafolio de servicios que le permite, mediante su afiliación empresarial e individual, generar la prestación de servicios que favorecen el desarrollo personal y laboral de sus afiliados y su grupo familiar.

La caja de compensación ofrece servicios como afiliaciones y subsidios, educación y cultura, recreación y turismo, vivienda y subsidio de vivienda, crédito social, alimentos y bebidas, salud y demás.

De acuerdo a lo dispuesto por la Ley 100 de 1993 “Se pretende que las cajas de compensación brinden a las personas y a las comunidades una mejor calidad de vida, mediante el cumplimiento progresivo de planes y programas que el Estado y la sociedad desarrollen para proporcionar la cobertura integral de las contingencias, especialmente las que menoscaban la salud y la capacidad económica de los habitantes del territorio nacional, con el fin de lograr el bienestar individual y la integración de la comunidad”.

Misión: Generar oportunidades para el cierre de brechas sociales.

Visión: Ser la empresa social de los colombianos

Al ser el departamento de mercadeo corporativo, un área transversal a todas las unidades de negocio, éste es el encargado de controlar y realizar las comunicaciones publicitarias de la caja de compensación.

Actualmente, en el departamento de mercadeo se realizan diversos envíos de información a los clientes, respecto a productos y/o servicios, que ofrecen las diferentes unidades de negocio; para

realizar dichos envíos, se segmenta la población por diferentes características, como puede ser la edad, género y otras más, según sea el mensaje que se quiere comunicar.

En el día, se puede realizar aproximadamente dos o tres o más envíos, según la necesidad, para ello se cuenta con la información de contacto y la respectiva autorización de manejo de datos personales.

Como el mercado actual, cada vez es más competitivo, se requiere que los procesos sean más ágiles y efectivos, por cuanto, un mensaje enviado a un dato errado genera pérdidas tanto de esfuerzo, como de oportunidad y posibles ventas.

Actualmente en el área solo se utilizan las herramientas de Access y Excel, aunque por el volumen creciente de información se está trabajando en un proyecto de Business Inteligencie, el cual dará opción a herramientas más avanzadas como SAP HANA y de Big data.

6 Problema

En el desarrollo normal del negocio, el departamento de mercadeo realiza varios envíos de publicidad, la cual, en varias ocasiones no es efectiva, debido a que rebotan algunos correos electrónicos enviados, al no ser válidos, o mensajes de texto, o envío de información física que no llega, porque la base de datos no está actualizada totalmente y en algunas ocasiones, demandas porque no se alcanzó a actualizar la información, con la revocación de la autorización de utilización de los datos personales, haciendo que no se enfoque los esfuerzos en clientes potenciales, siendo esto fundamental dado que las estrategias que se realizan en las diferentes unidades de negocio pierden su efectividad y posibles ventas.

6.1 Diagrama espina de pescado (causa – efecto)

DIAGRAMA DE ESPINA DE PESCADO (CAUSA-EFECTO)



Fuente: Elaboración propia

7 Justificación

Con este proyecto la empresa mejorará el contacto con los clientes, al tener la información de manera oportuna, depurada y confiable, lo que permitirá que las personas o empresas conozcan los productos y/o servicios que se prestan, igualmente, redundará en campañas más efectivas de mercadeo dirigidas a clientes potenciales identificados mediante el análisis y segmentación de los perfiles potenciales aumentando las ventas.

Adicionalmente, mediante el análisis de la información se disminuirán las quejas por contactar a personas que revocaron la autorización de manejo de sus datos personales o simplemente no lo autorizaron, evitando demandas millonarias que acarrearán costos innecesarios a la empresa.

7.1 Objetivo General

Optimizar los tiempos para la depuración de la base de datos que se suministran por diversas fuentes y que llegan al departamento mercadeo de la caja de compensación, con el fin de garantizar que la información que se genere semanalmente sea completa y actualizada, sirviendo de insumo

para el envío de información de promociones, productos y/o servicios a los clientes que han dado la autorización de Hábeas Data y que han pasado por un análisis mediante la aplicación de un modelo de segmentación de datos y que se perfilan como clientes potenciales.

8 Marco conceptual

Balanced scorecard: Son cuadros de mando integral que le permite a los negocios alinear los objetivos de los negocios con la estrategia de la empresa.

Bases de datos: Conjunto de información organizada la cual va a ser procesada por medio de un gestor de bases de datos, en este caso para realizar los diferentes envíos de publicidad a los clientes.

B.I.: Business Inteligencie es una herramienta bajo la cual una empresa soporta la toma de decisiones basándose en información con calidad y de forma oportuna generando conocimiento.

Campañas de mercado: Son las diferentes estrategias que realizan las unidades de negocio para poder contactar a los clientes y aumentar las ventas.

Cliente: Es toda persona natural o jurídica con la cual se tiene contacto, la cual puede ser afiliada o no a la Caja de Compensación.

Código de poblado: Es el número mediante el cual se identifica al departamento, la ciudad y el poblado en nuestro país, el cual es tomado de la división político-administrativa de Colombia definida por el DANE (Divipola).

Data warehouse: Es el proceso mediante el cual se extrae información de diferentes fuentes ya sean internas o externas, para que después que se les realiza la depuración, se almacenen en bases de datos consolidadas con el fin de que sean utilizadas por los negocios.

Depuración de la información: Proceso que se realiza a cada uno de los datos para poderlos dejar de forma consistente y utilizarlos en el contacto con el cliente.

Estandarización: Colocar la información uniformemente bajo unos conceptos y criterios específicos los cuales deben ser comunes en todos los registros.

ETL: Es el proceso de extraer, transformar y cargar información.

Fuentes de información: Cada uno de los insumos utilizados para alimentar las bases de datos, es decir los diferentes medios utilizados para la captura de los datos de contacto de las personas.

Hábeas Data: Es la ley implementada para la protección de datos personales de cada individuo, ésta fue creada por la ley 1581 de 2012 y reglamentada por el decreto 1377 de 2013.

Inconsistencias: Error en la captura de la información de números de identificación, tipos de documento, dirección, teléfono y correos electrónicos.

Macro: Es un programa diseñado para optimizar los tiempos en tareas repetitivas, siendo éste de fácil ejecución.

Minería de datos: Proceso mediante el cual se seleccionan, exploran, modifican o modelizan grandes cantidades de información para poder construir conocimiento o realizar predicciones.

Reportes: Es un documento que se le entrega a la unidad de negocio según los criterios solicitados.

Unidad de negocio: Es la unidad organizativa o área de la Caja de Compensación.

9 Estado del arte

Minería de datos y Big data

Las tecnologías de la información y Big Data están generando una inmensa transformación en el

mundo de los negocios, dando lugar a lo que se conoce como Nueva Economía o Economía Digital. Big Data se define en torno a 4 dimensiones, a saber: volumen, velocidad, variedad y veracidad, y puede verse como un conjunto de datos de gran volumen, que es dinámico, complejo y donde se puede obtener información de diversas fuentes y en una variedad de formatos. Las fuentes de Big Data incluyen, entre otros, datos de aplicaciones comerciales. El procesamiento de esta información requiere la aplicación de técnicas que permitan descubrir o extraer conocimiento, para que pueda ser utilizado en la toma de decisiones con el fin de encontrar nuevas oportunidades de negocio (al identificar las tendencias de los clientes), aumentar las ganancias, mejorar el resultado de las campañas publicitarias, realice operaciones en un tiempo más corto y logre clientes más satisfechos. (Jaramillo-Valbuena, 2018, Volumen 39)

“Business intelligence and analytics (BIA) trata sobre el desarrollo de tecnologías, sistemas, prácticas y aplicaciones para analizar datos críticos de negocios a fin de obtener nuevos conocimientos sobre negocios y mercados. Los nuevos conocimientos pueden utilizarse para mejorar los productos y servicios, lograr una mejor eficiencia operativa y fomentar las relaciones con los clientes. En este artículo, categorizaremos las actividades de investigación de BIA en tres direcciones de investigación amplias: (a) análisis de big data, (b) análisis de texto y (c) análisis de red. El artículo tiene como objetivo revisar las técnicas y los modelos más modernos y resumir su uso en las aplicaciones BIA.” (Ee-Peng Lim, Hsinchun Chen, and Guoqing Chen. 2013.)

“Las reglas de negocio guían el desempeño de una organización y su documentación ofrece una importante fuente de información para generar y mantener modelos tecnológicos, restricciones y otras estructuras en los sistemas de bases de datos. A pesar de su importancia, es muy difícil mantener la documentación actualizada de las reglas comerciales actuales que se implementan en los sistemas de información de una organización. Este documento describe la investigación sobre un método automático para el descubrimiento de reglas de negocio basado en técnicas de minería de procesos. El método propuesto extrae un conjunto de información relevante de los registros de ejecución de los sistemas de información y crea un conjunto de reglas de negocio actuales ('Tal como están') para el usuario.” (Raphael Crierie, Fernanda Araujo Baia o, y Flávia Maria Santoro. 2008)

Debido a la diversidad de métodos utilizados para ingresar información de autoría, la falta de estandarización de los datos bibliográficos se ha convertido en uno de los problemas que limitan el análisis de la información métrica en términos de tiempo de ejecución, confiabilidad de los indicadores y tamaño del cuerpo de los datos. El propósito del estudio fue proponer requisitos para mejorar la normalización de datos en el software de análisis métrico. Para lograr este objetivo, se hizo un diagnóstico de los principales métodos y técnicas utilizados en todo el mundo en este tipo de estudio. El resultado principal es la presentación de los requisitos que debe cumplir una aplicación para el procesamiento previo automático de datos. (Albo Hernández, 2018. pp. 55-73)

El hecho de no tener un método claro para ser eficiente en la búsqueda de decisiones en Business Intelligence (IN) genera altos costos en el uso de los recursos. Por lo tanto, la informática vino a resolver este problema por medio de una rama de sí mismo que es la inteligencia artificial y un subcampo como el aprendizaje automático (A-A). En el presente trabajo definiremos una metodología llamada M3S de IN, en la que se determina que un procedimiento apropiado a seguir tiene los mejores resultados en la extracción de datos, para obtener un mayor conocimiento y valor en la información. (Leonardo Morales, 2018. pp.1-7)

“La inteligencia comercial se está convirtiendo en un factor importante que puede ayudar a las organizaciones a administrar, desarrollar y comunicar sus bienes valiosos, como la información y el conocimiento. El objetivo principal de la minería de datos es la búsqueda de relaciones y patrones distintos que existen en los conjuntos de datos, pero están "ocultos" entre la gran cantidad de datos. La minería de datos se aplicó de manera efectiva en muchas áreas, entre ellas: control de inventario, diagnóstico de fallas, calificación crediticia, mercadeo, servicios financieros, programación, detección de fraudes y administración de redes. La minería de datos puede desempeñar un papel importante para la explotación de la inteligencia empresarial en el negocio electrónico.

Si bien las técnicas y herramientas para el comercio electrónico se están desarrollando activamente, hay mucho menos para la inteligencia de negocios electrónicos. Este documento contribuye al área de análisis de uso web para la inteligencia de negocios electrónicos al "casar" el negocio electrónico con la minería de datos, considerando la inteligencia de negocios como el

mejor entorno. Cuatro métricas, aplicadas de manera innovadora por primera vez en el campo del comercio electrónico, proporcionan capacidades para examinar datos sobre compras electrónicas en tiendas en línea y una visión detallada de la efectividad de los diferentes esfuerzos al responder preguntas comerciales relacionadas.” (Stavros Valsamidis, Ioannis Kazanidis, Sotirios Kontogiannis, and Alexandros Karakos. 2014.)

Actualmente podemos ver que hemos entrado en una nueva era de información en la que las compañías conocen de antemano nuestras preferencias y, de acuerdo con nuestro comportamiento en la red, pueden predecir qué productos preferiremos antes de su lanzamiento. Usan nuestra información para generar nuevas campañas de marketing con mayor seguridad de que los productos tendrán una mayor aceptación y al mismo tiempo reducirán los riesgos. © 2017 IEEE. (Judith, L.L. 2018. pp. 1-5)

En el desarrollo del presente proyecto y en el futuro de la Caja de compensación familiar es poder manejar estos datos e incluirlos en el Big data como pueden ser entre otros audios, para ello se ha mirado temas como el de este autor: “El discurso popular sobre Big data, que está dominado e influenciado por los esfuerzos de marketing de grandes desarrolladores de software y hardware, se centra en el análisis predictivo y los datos estructurados. Ignora el componente más grande de Big Data, que no está estructurado y está disponible como audio, imágenes, video y texto no estructurado. Se estima que los datos estructurados listos para análisis solo forman un pequeño subconjunto de Big data. Los datos no estructurados, especialmente los datos en formato de video son el componente más grande de big data que está parcialmente archivado aproximadamente un 95%” (Gandomi, A., Haider, M., 2014. p. 137)

“Durante la última década, la minería de procesos surgió como un nuevo campo de investigación que se centra en el análisis de procesos utilizando datos de eventos. Las técnicas clásicas de minería de datos, como la clasificación, agrupación, regresión, aprendizaje de reglas de asociación y minería de secuencias / episodios, no se centran en los modelos de procesos de negocios y, a menudo, solo se utilizan para analizar un paso específico en el proceso general. La minería de procesos se enfoca en los procesos de extremo a extremo y es posible debido a la creciente disponibilidad de datos de eventos y al nuevo proceso de descubrimiento y técnicas de verificación

de conformidad.

Los modelos de proceso se utilizan para el análisis (por ejemplo, simulación y verificación) y la promulgación por los sistemas BPM / WFM. Anteriormente, los modelos de procesos normalmente se hacían a mano sin utilizar datos de eventos. Sin embargo, las actividades ejecutadas por personas, máquinas y software dejan rastros en los llamados registros de eventos. Las técnicas de extracción de procesos utilizan dichos registros para descubrir, analizar y mejorar los procesos de negocios.

Recientemente, el Grupo de trabajo sobre minería de procesos publicó el Manifiesto de minería de procesos. Este manifiesto es apoyado por 53 organizaciones y 77 expertos en minería de procesos contribuyeron a él. La participación activa de usuarios finales, proveedores de herramientas, consultores, analistas e investigadores ilustra la creciente importancia de la minería de procesos como un puente entre la minería de datos y el modelado de procesos de negocios. La relevancia práctica de la minería de procesos y los interesantes desafíos científicos hacen de la minería de procesos uno de los temas "candentes" en Business Process Management (BPM).” (Wil van der Aalst. 2012)

Una parte muy importante en toda empresa que se maneje información y en especial de inteligencia de negocios es entregar información con calidad, integridad y que sea coherente, por tanto, un buen aporte es este documento que se cita a continuación: “El análisis de información asociado con la búsqueda de bases de datos, la minería y el análisis se puede ver como una capacidad de TI innovadora que puede mejorar el rendimiento de la empresa. Además, la experiencia favorable de una empresa (es decir, las percepciones de los beneficios) al utilizar datos de fuentes externas podría alentar la adquisición futura de análisis de Big data”. © 2014 Elsevier Ltd. (Kwon, O., Lee, N., Shin, B., 2014. p. 387)

En toda empresa donde se maneja información es necesario la minería de datos, por lo tanto, este artículo permite ahondar un poco más sobre el tema: “El avance de la tecnología para la gestión de bases de datos hace posible integrar diferentes tipos de datos, tales como imagen, video, texto, y otros datos numéricos, en una base de datos sencilla, facilitando el procesamiento multimedia.

Como resultado, la mezcla tradicional ad hoc de técnicas estadísticas y herramientas de gestión de datos no son adecuadas por más tiempo para analizar esta vasta colección de datos desiguales. La tecnología de Internet actual y su creciente demanda necesita el desarrollo de tecnologías de minería de datos más avanzadas para interpretar la información y el conocimiento de los datos distribuidos por todo el mundo. Es conocida la frase “los datos en bruto raramente son beneficiosos directamente”. Su verdadero valor se basa en: (a) la habilidad para extraer información útil la toma de decisiones o la exploración, y (b) la comprensión del fenómeno gobernante en la fuente de datos”. (Riquelme, Ruiz Roberto, Gilbert Karina, 2006. pp. 11 – 18)

La mayoría de las empresas tienen un sistema de información que soporta gran parte de las actividades diarias propias del sector de negocios en donde se esté desempeñando, este sistema puede ser sencillo o robusto todo depende de las exigencias del negocio, con el transcurso del tiempo estas aplicaciones llegan a tener la historia de la organización, los datos almacenados en las bases de datos pueden ser utilizados para argumentar la decisión que se quiera tomar.

El poder competitivo que puede tener una empresa se basa en la calidad y cantidad de la información que sea capaz de usar en la toma de decisiones; mediante la implementación de Inteligencia de Negocios se proporcionan las herramientas necesarias para aprovechar los datos almacenados en las bases de datos de los sistemas transaccionales para utilizar la información como respaldo a las decisiones, reduciendo el efecto negativo que puede traer consigo una mala determinación. (Rosado Gómez Alveiro Alonso, Rico Bautista Dewar Willmer, 2010)

El concepto de Inteligencia de negocios está asociado a la capacidad de un negocio de pensar, razonar y sentir. Pues eso es lo que se quiere obtener: que el razonamiento empresarial, en términos de toma de decisiones, sea efectivo y eficiente. Por lo tanto, lo visto con anterioridad es posible de lograr si se utiliza las últimas herramientas tecnológicas para la integración de sistemas de información, la transformación de datos en información y de información en conocimiento, y la explotación de sistemas gerenciales y de decisiones para un proceso de toma de resoluciones fundamentado. En consecuencia, la inteligencia de negocios es tal, siempre y cuando las decisiones gestionadas, en los niveles de la gerencia media y de la alta gerencia, provengan exclusivamente de las operaciones transaccionales de la empresa con la participación de mecanismos tecnológicos destinados para ese efecto. (Araníbar S. Juan Carlos, 2003)

Toda empresa necesita información para tomar decisiones y aún más si estamos usando datos de contacto para campañas de mercadeo. Business Inteligencie (BI) ayuda a las empresas a tomar decisiones basadas en un buen preanálisis y datos documentados, y permite que la información se presente cuando y donde las decisiones deben tomarse. La inteligencia de negocios en tiempo real (RTBI) presenta los números en tiempo real, brindando a los encargados de la toma de decisiones en las capas operativa y táctica con datos tan recientes como sea posible. (Darshan M. Tank, IJIEEB, vol.7, no.1, 2015, pp. 43-48)

“El mundo tradicional de las bases de datos relacionales y los almacenes de datos empresariales está siendo desafiado por el crecimiento en los volúmenes de datos, el aumento de datos no estructurados y semiestructurados, y el deseo de extraer información empresarial más valiosa. Para seguir siendo competitivos: estamos entrando en el mundo de los 'BIG DATA'. Están surgiendo soluciones basadas en hardware de gran escala basadas en el modelo de programación de reducción de mapas para procesamiento paralelo en hardware grande para abordar estos requisitos de BIG DATA que han desafiado las tecnologías tradicionales. El foco de esta charla está en el valor de negocio potencial que se creará en esta área al describir las oportunidades y los riesgos que surgen de la reciente aparición de la tecnología BIG DATA Analytics para empresas. El papel que pueden desempeñar las empresas en BIG DATA también se está discutiendo, y finalmente, la experiencia de Telefónica se explica al aplicar la tecnología BIG DATA, tanto internamente para mejorar sus propios procesos comerciales como externamente, donde estamos aplicando la tecnología para beneficiar directamente a nuestros clientes.” (V. Richard Benjamins. 2014.)

“Los correos electrónicos de empresa a consumidor (B2C) generalmente se generan al llenar los datos estructurados del usuario (por ejemplo, compras, eventos) en plantillas. La extracción de datos estructurados de correos electrónicos B2C permite a los usuarios realizar un seguimiento de información importante en varios dispositivos.

Sin embargo, también plantea varios desafíos, debido al requisito de tiempo de respuesta corto para el volumen masivo de datos, la diversidad y complejidad de las plantillas y las restricciones legales y de privacidad. En particular, los datos de correo electrónico son contenidos legalmente protegidos, lo que significa que nadie excepto el receptor puede revisar los mensajes o la

información derivada.

En este documento, primero introducimos un sistema que puede extraer información estructurada automáticamente sin requerir una revisión humana de ningún contenido personal. Luego nos centramos en cómo anotar los nombres de los productos de los textos extraídos, que es uno de los problemas más difíciles del sistema. Ni los métodos generales de aprendizaje, como los clasificadores binarios, ni los métodos de aprendizaje de estructura más específicos, como el Campo aleatorio condicional (CRF), pueden resolver bien este problema.

Para lograr esta tarea, se propone un enfoque híbrido, que básicamente entrena un modelo de CRF utilizando las etiquetas predichas por los clasificadores binarios (aprendizaje débil). Sin embargo, el rendimiento del aprendizaje débil puede ser bajo, por lo tanto, utilizamos el algoritmo de Maximización de Expectativas (EM) en CRF para eliminar el ruido y mejorar la precisión, sin la necesidad de etiquetar e inspeccionar correos electrónicos específicos. En nuestros experimentos, el modelo EM-CRF puede mejorar significativamente las anotaciones del nombre del producto sobre los aprendizajes débiles y los CRF simples.” (Weinan Zhang, Amr Ahmed, Jie Yang, Vanja Josifovski, y Alex J. Smola. 2015.)

Con relación a la depuración y organización de datos se han realizado múltiples trabajos e investigaciones en las que se pretende socializar las mejores prácticas para la limpieza de los datos que permitan la toma de decisiones basadas en datos certeros, dado a que en la actualidad se generan en grandes cantidades de información en su mayoría sin ningún criterio de organización.

En cuanto a la organización de datos, se puede tener en cuenta el artículo de URIBE, I., & JIMÉNEZ RAMÍREZ, C. (2009). HACIA UNA METODOLOGÍA PARA LA SELECCIÓN DE TÉCNICAS DE DEPURACIÓN DE DATOS en su abstrac se especifica lo siguiente: “Errores de digitación, datos inconsistentes, valores ausentes o duplicados, son algunos de los problemas que pueden presentar los datos almacenados en las bases y bodegas de datos, deteriorando su calidad y en consecuencia, la calidad de las decisiones que se tomen con base en el nuevo conocimiento obtenido a partir de ellos. Este artículo pone de manifiesto la necesidad de una guía metodológica que apoye a los analistas de datos en la selección de las técnicas de depuración, considerando los diferentes tipos de errores en los datos y la naturaleza de estos”.

También se tiene diversidad de trabajos realizados con la identificación de afectaciones en las bases de datos, es el caso de [Rahm y Do, 2000] en el que se identifican los problemas de la calidad de los datos mediante la limpieza de los datos, [Kim et. al., 2003] en este trabajo se determina que la limpieza de datos es una parte integral del procesamiento de datos para generar utilidad de los mismos, se clasifican las diversas incoherencias que deben eliminarse en el procesamiento de las bases de datos y se dan los pasos para depurar los datos.

Oliveira et. al. [Oliveira et.al., 2005] no sólo realizan una taxonomía con treinta y cinco problemas de calidad de los datos, sino que plantean métodos semiautomáticos para detectarlos, los cuales representan mediante árboles binarios. Los árboles corresponden al razonamiento que se necesita hacer para detectar un problema particular.

El tema de los datos duplicados es uno de los más recurrentes al momento de realizar la depuración de los datos [Elmagarmid et. al., 2007], el cual tiene como meta identificar registros que se refieren a la misma entidad del mundo real, aun si los datos no son idénticos, esto es, se trata de la detección de atributos o registros que tienen contenidos distintos pero debieran ser el mismo. Una misma entidad del mundo real puede aparecer representada dos o más veces (duplicada) a través de una o varias bases de datos, en tuplas con igual estructura, que no comparten un identificador único y presentan diferencias textuales en sus valores. En el capítulo 4 se trata este tema en detalle.

[Medina y Galván, 2007] busca dar solución a los datos faltantes.

Como se puede apreciar existen varias investigaciones que buscan identificar y solucionar los temas de limpieza y depuración de las bases de datos para contar con información totalmente confiable que permita la toma de decisiones con base a datos totalmente fidedignos.

10 Objetivos Específicos, actividades y cronograma

Objetivo Específico No. 1	
Elaborar el manual para el manejo adecuado de la información.	
Metodología	Se creará el manual para el manejo adecuado de la información y los estándares para la homologación los datos.

Alcance	En el manual se especificarán los estándares que se deben tener en cuenta para las homologaciones de tipo de documento, direcciones, teléfonos y correos electrónicos.
Actividades	Definir el estándar que se va a utilizar para abreviar el tipo de documento, que llega en la codificación de las diferentes fuentes. Colocar el estándar utilizado en la nomenclatura DANE Anexar la tabla Divipola para la estandarización de la información. Describir los criterios que se deben tener en cuenta en la depuración de teléfonos y correos electrónicos.
Cronograma	Un mes después de aprobado el proyecto

Cronograma Primer Objetivo Específico							
No	Nombre	Duración Días hábiles	Comienzo	Fin	Entregable	Responsable	Observaciones
1	Elaborar el manual para el manejo adecuado de la información.	1 MES	1-mar-19	29-mar-19			
1.1	Recopilación de la información - tipo de documento, direcciones, teléfonos y correos electrónicos.	10 días	1-mar-19	14-mar-19		Caja de Compensación	La caja de compensación suministrara al proyecto la información de sus bases de datos.
1.2	Definir el estándar que se va a utilizar para abreviar el tipo de documento.	2 días	15-mar-19	18-mar-19		Proyecto	
1.3	Colocar el estándar utilizado en la nomenclatura DANE	3 días	19-mar-19	21-mar-19		Proyecto	
1.4	Anexar la tabla Divipola para la estandarización de la información.	3 días	22-mar-19	26-mar-19		Proyecto	
1.5	Describir los criterios que se deben tener en cuenta en la depuración de teléfonos y correos electrónicos.	3 días	27-mar-19	29-mar-19	Manual	Proyecto	Finalizado el análisis de la información se tiene como entregable el Manual con los estándares para el manejo adecuado de la información.

Objetivo Específico No. 2	
Crear las macros y plantillas de depuración y consolidación de la información.	
Metodología	Se implementarán macros independientes para procesar los datos de direcciones, teléfonos, correos electrónicos. De igual forma se elaborarán las macros con el fin de dejar el registro más actualizado tanto en las fuentes como en la base de datos.
Alcance	Se crearán las macros para dar cumplimiento al proyecto, utilizando el lenguaje de Visual Basic.
Actividades	Creación de plantillas en Excel para facilitar la depuración de información. Creación de archivos en Excel, los cuales se encargarán de consolidar la información de las múltiples fuentes, dejando el registro más actualizado.

	Creación de macros en Access para la actualización de las bases de datos, de igual forma permitirá la generación de los insumos para la entrega de requerimientos solicitados por las unidades de negocio.
Cronograma	Cinco meses después de aprobado el manual

Cronograma Segundo Objetivo Especifico							
No	Nombre	Duración Días hábiles	Comienzo	Fin	Entregable	Responsable	Observaciones
2	Crear las macros y plantillas de depuración y consolidación de la información.	5 MESES	1-abr-19	11-sep-19			
2.1	Creación de plantillas en Excel para facilitar la depuración de información.	15 días	1-abr-19	23-abr-19	Plantillas de Depuración	Proyecto	
2.2	Creación de archivos en Excel, los cuales se encargarán de consolidar la información de las múltiples fuentes, dejando el registro más actualizado.	45 días	24-abr-19	28-jun-19	Consolidado de la información en Excel	Proyecto	
2.3	Creación de macros en Access para la actualización de las bases de datos, de igual forma permitirá la generación de los insumos para la entrega de requerimientos solicitados por las unidades de negocio.	42 días	2-jul-19	11-sep-19	Macros	Proyecto	

Objetivo Específico No. 3	
Generar los formatos para el envío de correos electrónicos y mensajes de texto para los clientes que hayan autorizado el manejo de información cumpliendo con la ley de habeas data.	
Metodología	Se diseñarán las plantillas con la información necesaria que permita al área encargada enviar la información según la segmentación que requiera.
Alcance	Las platillas serán creadas en Excel ó en archivos planos.
Actividades	Se crearán las plantillas de acuerdo a la información requerida. Se elaborarán carpetas compartidas de acceso exclusivo para mantener la seguridad de la información. Se dejarán los archivos en las rutas compartidas previamente definidas.
Cronograma	Un mes después de creados los desarrollos y realizadas las pruebas.

Cronograma Tercer Objetivo Especifico							
No	Nombre	Duración Dias hábiles	Comienzo	Fin	Entregable	Responsable	Observaciones
3	Generar los formatos para el envío de correos electrónicos y mensajes de texto para los clientes que hayan autorizado el manejo de información cumpliendo con la ley de habeas data	1 MESES	12-sep-19	11-oct-19			
3.1	Se crearán las plantillas de acuerdo a la información requerida.	5 días	12-sep-19	18-sep-19	Plantillas	Proyecto	
3.2	Se elaborarán carpetas compartidas	12 días	19-sep-19	4-oct-19	Carpetas compartidas	Proyecto	Las carpetas compartidas son de acceso exclusivo para mantener la seguridad de la información.
3.3	Se dejarán los archivos en las rutas compartidas previamente definidas.	5 días	7-oct-19	11-oct-19	Compartir la información para la depuración	Proyecto	Cada área involucrada deberá implementar la metodología descrita en el manual para depurar la información con las herramientas implementadas.

Objetivo Específico No. 4	
Crear un modelo de segmentación adecuado para los clientes potenciales de envío de información publicitaria	
Metodología	Por medio del aplicativo K-nime se realizará una segmentación y modelación de datos con la información ya depurada para perfilar los clientes potenciales que se encuentran con capacidad de adquirir los servicios de la caja de compensación.
Alcance	Creación de un modelo inicial con la información existente en el momento.
Actividades	Selección de datos: el análisis se realizará para una muestra de 200.000 registros. Limpieza de datos: selección de los datos que se van a tener en cuenta. Modelado: construcción del modelo Evaluación: evaluación del modelo y su funcionalidad.
Cronograma	Veinte días después de depurada la información.

Cronograma Cuarto Objetivo Especifico							
No	Nombre	Duración Días hábiles	Comienzo	Fin	Entregable	Responsable	Observaciones
4	Generar los formatos para el envío de correos electrónicos y mensajes de texto para los clientes que hayan autorizado el manejo de información cumpliendo con la ley de habeas data	20 días	12-oct-19	8-nov-19			
4.1	Selección de datos	5 días	12-oct-19	18-oct-19		Proyecto	
4.2	Limpieza de datos	5 días	19-oct-19	24-oct-19		Proyecto	
4.3	Modelado	5 días	25-oct-19	31-oct-19	Selección del mejor modelo	Proyecto	
4.4	Evaluación del modelo	5 días	1-nov-19	8-nov-19	Carpetas compartidas	Proyecto	Evaluación e Implementación del modelo de perfilamiento

10.1 Metodología

Para dar alcance a la propuesta, se realizará la documentación de la estandarización de la información, luego se realizarán las plantillas y las macros que permitirán la depuración de la información, para que posteriormente, sea cargada en la base de datos, asegurando que quede el registro más actualizado, disminuyendo tiempo y reproceso, que se generan antes de la implementación del proyecto.

Toda la información tanto de contacto como demográfica de los afiliados será suministrada por la Caja de compensación.

A lo largo de este proceso, se estará garantizando que la autorización o negación que dé el usuario se actualice en el menor tiempo posible, de igual forma se reducirá el tiempo para solucionar los derechos de petición que envíe la Superintendencia respecto a los casos de Hábeas Data.

Finalmente se realizará un modelo de segmentación que permita identificar a que grupo de personas se les debe enviar un tipo de información determinada según el portafolio que esté vigente en el momento del proyecto.

10.2 Presupuesto General del Proyecto

Presupuesto aproximado en miles de pesos				
Rubro	Valor unitario	Financiación propia	Otra institución (indique nombre(s))	Total
Personal	\$ 3.000.000		Caja de compensación Familiar	\$ 27.000.000
Equipos	\$ 10.402.099		Caja de compensación Familiar	\$ 10.402.099
Software	\$ 30.660		Caja de compensación Familiar	\$ 328.369
Total	\$ 13.432.759			\$ 37.730.468

11 Viabilidad financiera

Nos da una viabilidad económica de la propuesta a través de la relación costo beneficio. El presupuesto para el proyecto es de \$37.730.468.

El proyecto es viable porque reduciría el tiempo de depuración de la información de 3 semanas a 3 días, con posibilidad de seguir disminuyendo el tiempo a medida que se retroalimente el sistema.

12 Plan de actividades

No.	Actividad	Mes 1				Mes 2				Mes 3				Mes 4				Mes 5				Mes 6				Mes 7				Mes 8				Mes 9			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.1	Recopilación de la información - tipo de documento, direcciones, teléfonos y correos electrónicos.																																				
1.2	Definir el estándar que se va a utilizar para abreviar el tipo de documento.																																				
1.3	Colocar el estándar utilizado en la nomenclatura DANE																																				
1.4	Anexar la tabla Divipola para la estandarización de la información.																																				
1.5	Describir los criterios que se deben tener en cuenta en la depuración de teléfonos y correos electrónicos.																																				
2.1	Creación de plantillas en Excel para facilitar la depuración de información.																																				
2.2	Creación de archivos en Excel, los cuales se encargarán de consolidar la información de las múltiples fuentes, dejando el registro más actualizado.																																				
2.3	Creación de macros en Access para la actualización de las bases de datos, de igual forma permitirá la generación de los insumos para la entrega de requerimientos solicitados por las unidades de negocio.																																				
3.1	Se crearán las plantillas de acuerdo a la información requerida.																																				
3.2	Se elaborarán carpetas compartidas																																				
3.3	Se dejarán los archivos en las rutas compartidas previamente definidas.																																				
4.1	Selección de datos																																				
4.2	Limpieza de datos																																				
4.3	Modelado																																				
4.4	Evaluación del modelo																																				
	Retroalimentación del proyecto																																				

13 Plan de adquisiciones

13.1 Necesidad de adquisiciones externas

Una de las principales actividades para la ejecución del proyecto, es la búsqueda de proveedores, que suministrarán los productos y servicios requeridos. Las adquisiciones se realizarán desde el inicio del proyecto

Los responsables de realizar las labores de compra y contratación son:

Gerente del proyecto: aprueba las contrataciones.

Especialista de costos: ejecuta las labores de compra y adquisiciones.

13.2 Identificación de adquisiciones

Rubro	Descripción	Valor total	Uso de la adquisición
-------	-------------	-------------	-----------------------

Desarrollador	Trabajo de desarrollo	\$ 27.000.000	Parametrizaciones, creación de macros y base de datos del proyecto de B.I.
Computador	Workstation HP Z240 3.6 GHz Intel Xeon E3 v5 E3-1270V5 Black Tower	\$ 10.402.099	Usado para todo el proceso y los desarrollos
Licencia Office 365	Licencia de Office que incluye Access	\$ 328.369	Usado para todo el proceso y los desarrollos
Total		\$ 37.730.468	

13.3 Tipos de contrato a utilizar

Producto o servicio a adquirir	Tipo de Contrato	Forma de contactar proveedores	Responsable de la compra
Desarrollador	Por cumplimiento de objetivos	Pool de proveedores	Project Manager y especialista de costos
Computador	Contrato de precio fijo	Pool de proveedores	Project Manager y especialista de costos
Licencia Office 365	Contrato de precio fijo	Pool de proveedores	Project Manager y especialista de costos

14 Plan de riesgos

ROLES Y RESPONSABILIDADES DE GESTIÓN DE RIESGOS			
PROCESO	ROLES	PERSONAS	RESPONSABILIDADES
Planificación de Gestión de los Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Apoyo definiciones y toma decisiones Ejecutar Actividades
Identificación de Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Proveer definiciones Ejecutar Actividades
Análisis Cualitativo de Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Proveer definiciones Ejecutar Actividades para mitigar
Análisis Cuantitativo de Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Proveer definiciones Ejecutar Actividades para mitigar
Planificación de Respuesta a los Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Proveer definiciones Ejecutar Actividades para mitigar
Seguimiento y Control del Riesgos	Equipo de Gestión de Riesgos Lider Apoyo Miembros	Project manager Sponsor Equipo Proyecto	Dirigir actividad, responsable directo Proveer definiciones Ejecutar Actividades

14.1 Valoración Cualitativa y cuantitativa de riesgos

Escala de probabilidad

PROBABILIDAD	0,9	MUY ALTA	Sucede en proyectos similares varias veces al año
	0,7	ALTA	Ha ocurrido en proyectos similares en el último año (o de probable ocurrencia en un lapso de 1 a 5 años)
	0,5	MEDIA	Ha ocurrido en proyectos similares en los últimos 10 años
	0,3	BAJA	Ha ocurrido en proyectos similares en un periodos superior a 10 años
	0,1	MUY BAJA	No ha ocurrido en otros proyectos similares

Escala de impacto

	IMPACTO				
OBJETIVO DEL PROYECTO	MUY BAJO	BAJO	MEDIO	ALTO	MUY ALTO

COSTO	Incremento de Costo insignificante	Incremento de Costo <10%	Incremento de Costo de 10% a 25%	Incremento de Costo de 25% a 50%	Incremento de Costo > 50%
TIEMPO	Desviación de Schedule insignificante	Desviación de Schedule < 5%	Desviación de Schedule 5% a 15%	Desviación de Schedule 15% a 25%	Desviación de Schedule > 25%
ALCANCE	Desviación poco notoria de Alcance	Áreas menores de Alcance afectadas	Principales áreas de Alcance afectadas	Cambio de Alcance inaceptable para el cliente	Producto final del proyecto no es utilizable
CALIDAD	Degradación poco notoria de calidad	Afectadas solo aplicaciones muy exigentes	La reducción de Calidad requiere aprobación del Cliente	La reducción de Calidad es inaceptable para el Cliente	Producto final del proyecto no es utilizable

Matriz de probabilidad por impacto

			IMPACTO AMENAZAS					IMPACTO OPORTUNIDADES				
			0,05	0,1	0,2	0,4	0,8	0,8	0,4	0,2	0,1	0,05
			MUY BAJO	BAJO	MEDIO	ALTO	MUY ALTO	MUY ALTO	ALTO	MEDIO	BAJO	MUY BAJO
PROBABILIDAD	0,9	MUY ALTA	0,045	0,09	0,18	0,36	0,72	0,72	0,36	0,18	0,09	0,045
	0,7	ALTA	0,035	0,07	0,14	0,28	0,56	0,56	0,28	0,14	0,07	0,035
	0,5	MEDIA	0,025	0,05	0,1	0,2	0,4	0,4	0,2	0,1	0,05	0,025
	0,3	BAJA	0,015	0,03	0,06	0,12	0,24	0,24	0,12	0,06	0,03	0,015
	0,1	MUY BAJA	0,005	0,01	0,02	0,04	0,08	0,08	0,04	0,02	0,01	0,005

14.2 Monitoreo y seguimiento

Monitoreo de Riesgos	Seguimiento y Gestión de Riesgos	Plan de respuesta a los riesgos
-----------------------------	---	--

1. Registro de riesgos, validación de los supuestos del proyecto	1. Realizar auditoria mensual de los riesgos identificados	1. Con base en la priorización de los riesgos Negativos gestionar si se deben evitar, transferir, mitigar o aceptar. Para el caso de las oportunidades evaluar si se deben explorar, mejorar, compartir o aceptar
2. Plan de dirección del proyecto (Documentación de matriz de riesgos aplicada al proyecto)	2. Reevaluación periódica de los riesgos	2. Enfocarse primero en los riesgos de mayor valoración de acuerdo a matriz de probabilidad por impacto
3. Informes de desempeño y avance del proyecto	3. Seguimiento a desviaciones de líneas base de Alcance, Cronograma y Costos (Análisis de variación y tendencias)	3. Evaluar planes de mitigación enfocados en reducir la probabilidad de ocurrencia: Planes de choque con mayores recurso de personal, actividades en el presupuesto y cronograma según las necesidades del proyecto
4. Líneas Base de Alcance, Cronograma y Costos	4. Plan de utilización de Reservas	4. Acciones requeridas para gestionar cada riesgo (con asignación de responsables por cada acción)

14.3 Matriz de riesgos aplicada al proyecto

No.	Causa	Evento	Consecuencia	Categoría	Responsable	Probabilidad	Impacto	Probabilidad x Impacto (Nivel de Riesgo)	Plan de Respuesta
R-001	Si la recopilación de la información no se reúne en dos semanas	Inicio tardío de construcción del proyecto	Impacto en el cronograma y promesa de valor del proyecto	internos	Encargado de bases de datos	Medio (0,5)	Alto (0,4)	Alto (0,2)	1. Priorizar la hoja de especificaciones de la información 2. Seguimiento de la entrega de la información
R-002	Definición de estándares	Desviaciones de los estándares	Impacto negativo a las base de datos, Alcance, costos y cronograma	Internos	Project	Medio (0,5)	Medio (0,2)	Medio (0,1)	1. Incluir desde el inicio del proyecto todos los interesados 2. Seguimiento y gestión de interesados del proyecto 3. Proceso de

									control de cambios
R-003	Elaboración de los criterios en la depuración de teléfonos y correos	Se podría demorar más la ejecución	Impacto negativo a las líneas base de Alcance, costos y cronograma	Interno	Project Manager	Medio (0,5)	Medio (0,2)	Medio (0,1)	1. Monitoreo y control de línea base de cronograma 2. Seguimiento y gestión de criterios
R-004	Creación de plantillas y archivos en Excel	Daño del equipo Se tendría un mayor alcance por Estandarización de versiones de Excel	Impacto negativo a las líneas base de Alcance, costos y cronograma	Interno	Ing. Proceso	Bajo (0,3)	Alto (0,4)	Medio (0,12)	1. Elaborar plantillas y copias de respaldo 2. Generación de archivos
R-005	Creación de macros en Access	Se podría impactar los tiempos de creación de macros	Impacto negativo a las líneas base de Alcance, costos y cronograma	Internos	Project Manager	Medio (0,5)	Medio (0,2)	Medio (0,1)	1. Plan de gestión de recursos 2. Monitoreo de avances en etapa de construcción del proyecto
R-006	Generación de rutas y carpetas compartidas	Retraso en actividades del proyecto	Impacto negativo a las líneas base de costos y cronograma	Organizacionales	Project Manager	Baja (0,3)	Alto (0,4)	Medio (0,12)	1. Plan de gestión de recursos para aseguramiento del personal 2. Uso de contratos marcos en caso de requerirse suministro de personal externo
R-007	Selección de datos, limpieza, modelado	Selección errada de la información de datos	Mayor agilidad en la entrega de la información, promesa de valor	Internos	Sponsor	Medio (0,5)	Medio (0,2)	Medio (0,1)	1. sondeo de datos 2. Gestión de modelado del proyecto

15 Plan de interesados

15.1 Registro de interesados

Nombre	Cargo	Departamento / División	Intereses	Etapas de mayor Influencia en el Proyecto
--------	-------	-------------------------	-----------	---

Jefe mercadeo	Jefe de departamento mercadeo	Mercadeo	Económico (Ganancias por incremento en ventas)	Todo el proyecto
Jefe UES	Jefes de las unidades de servicios	Las diferentes unidades de servicios	Económico (Ganancias por incremento en ventas)	Todo el proyecto
Jefe de información	Jefe del área de información	Mercadeo	Asegurar el cumplimiento de alcance y especificaciones técnicas de acuerdo a estándares	Todo el proyecto
Jefe financiero	Jefe del área financiera	Mercadeo	Es el Sponsor del proyecto, espera cumplimiento de líneas bases y promesa de valor del proyecto	Todo el proyecto

15.2 Gestión de interesados

Interesado	Tipo de interesado	Nivel de interés	Nivel de influencia	Acciones posibles de parte del interesado tanto de impacto positivo como de impacto negativo	Estrategias para la gestión del interesado	Qué tipo de información y con qué frecuencia se enviará esta al interesado para mantenerlo informado sobre la ejecución del proyecto
Jefe mercadeo	Interno (Cliente)	Líder	Alto	Solicitud de cronograma más corto	Involucramiento y aprobación de líneas base	Informes de avance del proyecto (Mensual)
Jefe UES	Interno (Cliente)	Líder	Alto	Solicitud de cronograma más corto	Involucramiento y aprobación de líneas base	Informes de avance del proyecto (Mensual)
Jefe de información	Interno (Cliente)	Líder	Alto	Solicitud de inclusión de adicionales que no estén alineados con los objetivos del proyecto	Involucramiento y aprobación de líneas base	Informes de avance del proyecto (Mensual)
Jefe financiero	Interno (Sponsor)	Líder	Alto	Solicitud de cronograma más corto	Involucramiento y aprobación de líneas base	Informes de avance del proyecto (Mensual)

16 Conclusiones y recomendaciones

La información que llega de diversas fuentes a la caja de compensación debe ser limpiada, depurada, clasificada y segmentada para que los esfuerzos de envío de información del área de mercadeo a las personas naturales y a las empresas sean efectivos y así aumentar las ventas.

Se requiere que la aplicación de inteligencia de negocios mediante el empleo de métodos sistematizados que permitan utilizar información con calidad, integridad y que sea coherente para los procesos de captación de clientes para la caja de compensación.

La implementación de la metodología y la utilización de las herramientas planteadas en el presente trabajo permitirán la realización de la labor de envío de información de una manera más rápida, efectiva, sin perjuicio para las personas y para la caja de acuerdo a la aplicación de la Ley de Habeas Data.

Se sugiere aplicar la metodología planteada en el presente proyecto, socializarla con las diferentes áreas y personas que intervienen en el proceso para que se apropien de su funcionamiento.

Se recomienda que la presente metodología haga parte de los procesos de calidad de la caja de compensación familiar así se asegurará que los empleados la utilicen de forma adecuada.

Cada vez que ingrese información nueva de diversas fuentes debe ser tratada con la metodología planteada antes de ser utilizada para el envío de información a los posibles clientes.

17 Bibliografía

Albo Hernandez, Ramón Orlando et al. Requerimientos para mejorar la normalización de datos en software de análisis métricos de la información. Rev. cuba. inf. cienc. salud [online]. 2018, vol.29, n.1, pp. 55-73.

<https://www-scopus-com.loginbiblio.poligran.edu.co/record/display.uri?eid=2-s2.0-85047161275&origin=resultslist&sort=plf-f&src=s&st1=data+mining&nlo=&nlr=&nls=&sid=484523cda3846810555fd18de9a18d92&sot=b&sdt=cl&cluster=scolang%2c%22Spanish%22%2ct&sl=26&s=TITLE-ABS->

[KEY%28data+mining%29&relpos=21&citeCnt=2&searchTerm=](#)

Ee-Peng Lim, Hsinchun Chen, y Guoqing Chen. 2013. Business Intelligence and Analytics: dirección de investigación. Transacciones de ACM en sistemas de información gerencial. Volumen 3, número 4, Artículo 17 (Enero 2013), 10 páginas.
<http://dx.doi.org.loginbiblio.poligran.edu.co:2048/10.1145/2407740.2407741>

Jaramillo-Valbuena, 2018, Volumen 39. Minería sobre streams de datos
<https://www-scopus-com.loginbiblio.poligran.edu.co/record/display.uri?eid=2-s2.0-85044729428&origin=resultslist&sort=plf-f&src=s&st1=data+mining&nlo=&nlr=&nls=&sid=484523cda3846810555fd18de9a18d92&sot=b&sdt=cl&cluster=scolang%2c%22Spanish%22%2ct&sl=26&s=TITLE-ABS-KEY%28data+mining%29&relpos=20&citeCnt=0&searchTerm=>

Leonardo Morales, 2018. pp.1-7. Metodología para procesos de inteligencia de negocios con mejoras en la extracción y transformación de fuentes de datos
<https://www-scopus-com.loginbiblio.poligran.edu.co/record/display.uri?eid=2-s2.0-85049894081&origin=resultslist&zone=contextBox>

Algoritmos de aprendizaje automático para análisis y predicción de datos. Judith, L.L. 2017 IEEE 37a Convención de América Central y Panamá, CONCAPAN 2017
Volumen 2018, 31 Enero 2018, páginas 1-5
<https://www-scopus-com.loginbiblio.poligran.edu.co/record/display.uri?eid=2-s2.0-85050562822&origin=resultslist&zone=contextBox>

Raphael Cerie, Fernanda Araujo Baiao , y Flávia Maria Santoro. 2008. Identificación de reglas de negocio utilizando minería de procesos. En actas acompañantes del XIV simposio Brasileño en multimedia y web. ACM, Nueva York, NY, USA, 241-246.
<http://dx.doi.org.loginbiblio.poligran.edu.co:2048/10.1145/1809980.1810046>

Rosado Gomez Alveiro Alonso, 2010, Ciencia y tecnología 2010, XVI (44). Inteligencia de

negocios <https://www.redalyc.org/articulo.oa?id=84917316060>

Stavros Valsamidis, Ioannis Kazanidis, Sotirios Kontogiannis, y Alexandros Karakos. 2014. Una metodología propuesta para la medición de la inteligencia de negocios electrónicos utilizando técnicas de minería de datos. En Actas de la 18a Conferencia Panhelénica sobre Informática (PCI '14). ACM, Nueva York, NY, EE. UU., Artículo 3, 6 páginas.

<https://doi-org.loginbiblio.poligran.edu.co/10.1145/2645791.2645828>

Uribe, I., & Jiménez Ramírez, C. (2009). Hacia una metodología para la selección de técnicas de depuración de datos (Junio 2009). Avances en Sistemas e Informática, 6(1), 185-190.

<https://revistas.unal.edu.co/index.php/avances/article/view/14485/15374>

V. Richard Benjamins. 2014. Big Data: de exageración a la realidad?. En los procedimientos de la 4th conferencia internacional sobre inteligencia en la Web, minería y semántica (WIMS14) (WIMS '14). ACM, Nueva York, NY, USA, Artículo 2, 2 páginas.

<https://doi-org.loginbiblio.poligran.edu.co/10.1145/2611040.2611042>

Weinan Zhang, Amr Ahmed, Jie Yang, Vanja Josifovski, y Alex J. Smola. 2015. Extracción de información de productos de correos electrónicos. En las actas de la 21ª Conferencia Internacional SIGMD de ACM sobre Descubrimiento de Conocimientos y Minería de Datos (KDD '15). ACM, Nueva York, NY, EE. UU., 2257-2266.

<https://doi-org.loginbiblio.poligran.edu.co/10.1145/2783258.2788580>

Wil van der Aalst. 2012. Proceso de minería: Resumen y oportunidades. Transacciones de ACM en sistemas de información gerencial. Volumen 3, número 2, Artículo 7 (Julio 2012), 17 páginas.

<http://dx.doi.org.loginbiblio.poligran.edu.co/2048/10.1145/2229156.2229157>