



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Desarrollos de Proyectos en el Sector Bancario

INFORME DE ACTIVIDADES PROFESIONALES

Que para obtener el título de

Ingeniera en Computación

P R E S E N T A

Catalina Martínez Martínez

ASESORA DE INFORME

Dra. María del Pilar Ángeles



Ciudad Universitaria, Cd. Mx., 2019

INDICE

Introducción.....	3
Capítulo 1 – Empresa.....	5
1.1 Antecedentes.....	5
1.2 Estructura del modelo de negocios.....	5
1.3 Historia.....	5
1.4 Organigrama.....	7
Capítulo 2 –Metodología y Descripción de Proyectos.....	9
2.1 Metodología Corporativa.....	9
2.1.1 Gestión de Iniciativas.....	11
2.1.2 Desarrollo.....	12
2.1.3 Gestión del Servicio.....	13
2.2 Proyecto 1 – Generación Modelo de Datos.....	15
2.2.1 Antecedentes.....	15
2.2.2 Objetivo.....	15
2.2.3 Descripción.....	16
2.2.4 Solución.....	16
2.2.5 Resultados.....	19
2.3 Proyecto 2 – Cédula Comparativa.....	21
2.3.1 Antecedentes.....	21
2.3.2 Objetivo.....	22
2.3.3 Descripción.....	22
2.3.4 Solución.....	23
2.3.5 Resultados.....	27
2.4 Proyecto 3 - Hermes Global.....	29
2.4.1 Antecedentes.....	29
2.4.2 Objetivo.....	30
2.4.3 Descripción.....	30
2.4.4 Solución.....	32
2.4.5 Resultados.....	42
Capítulo 3 – Marco Teórico Minería de Datos.....	47
3.1 ¿Qué es la Minería de Datos?.....	48
3.2 ¿Cómo funciona la Minería de Datos?.....	48

3.2.1 Modelado Descriptivo	49
3.2.2 Modelado Predictivo	50
3.2.3 Modelado Prescriptivo	50
3.3 Ciclo de la Minería de Datos	51
3.3.1 Definir Problema.	52
3.3.2 Preparar Datos	53
3.3.3 Explorar datos	54
3.3.4 Generar Modelo	55
3.3.5 Explorar y validar los modelos	57
3.3.6 Implementar y actualizar los modelos.....	59
Capítulo 4 – Modelo de Minería	61
4.1 Situación actual	61
4.2 Objetivo	62
4.3 Desarrollo de la solución.....	62
4.3.1 Creación del Modelo	63
4.3.2 Implantación del Modelo de Minería	71
4.5 Resultados	76
Capítulo 5 - Conclusiones	77
Glosario.....	78
Referencias:	79

Introducción

La obtención del grado de licenciatura constituye una parte fundamental de la formación académica, considerando los factores económicos, sociales, el tiempo que ha transcurrido desde el término de mis estudios y las diferentes opciones de titulación que ahora permite la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, he concluido que la mejor opción de titulación debe ser por experiencia profesional.

Este trabajo describe mis actividades profesionales, donde los conocimientos de ingeniería en computación adquiridos en la Facultad de Ingeniería fueron la base para mi desarrollo profesional en la empresa del sector financiero.

El informe se centra en los proyectos de software, actividades de operación y soporte relevantes realizados durante mi carrera en el sector bancario, mi participación consistió en asistir, analizar, programar, liderar, así como la gestión de proyectos de software, bajo los procedimientos establecidos, metodologías y guías implementadas en la empresa, para el cumplimiento de los objetivos estratégicos de la organización, el mejoramiento de los procesos, mantenimiento para la reducción de costos operativos y tiempo en los sistemas de información.

El objetivo principal es proporcionar una descripción detallada de mi trabajo profesional desarrollado en el proyecto de Modelo de Minería “Detección de Empresas con Potencial de Crecimiento en Saldo de Captación”. El trabajo descrito hace uso de algunos aspectos teóricos, documentación (diagramas, modelos de funcionamiento, arquitectura de sistemas, flujos y secuencia de componentes) e información producida durante las fases de un proyecto (inicio, análisis, diseño, construcción, pruebas e implantación), así como el soporte y mantenimiento a la aplicación.

El reporte también describe mi participación en tres proyectos más, que fueron desarrollados antes del proyecto principal, estos fueron seleccionados de diferentes periodos y en los cuales contaba con menos experiencia, lo anterior con el fin de mostrar mi avance en la adquisición de la misma y mostrar mi crecimiento profesional.

Mi participación en los diferentes proyectos fue desde programador hasta líder de proyecto, realizando actividades de análisis del requerimiento (entendimiento de la necesidad del negocio, identificando oportunidades de mejora o extensión del alcance inicial), diseñando soluciones, análisis y plan de riesgos, a la par de la gestión de los requerimientos a las áreas de infraestructura, seguridad, almacenamiento, entre otras. Trabajando con una diversidad de personas de sistemas, de las diferentes áreas de negocio y diferentes niveles que me enseñaron nuevas visiones de los proyectos, alcances, así como la importancia de las estrategias de la organización.

En concreto, este reporte tiene como objetivo mostrar que mi labor principal como ingeniero en computación es diseñar sistemas de información que generen valor a las organizaciones, lo cual se traduce en incrementar las ganancias y/o bajar los costos y tiempos de procesamiento, utilizando la infraestructura y herramientas adquiridas por la organización o empresa.

Capítulo 1 – Empresa.

1.1 Antecedentes.

La empresa es una institución financiera privada con importante presencia en México que ofrece una amplia variedad de productos y servicios financieros, es la institución líder en México en cartera de crédito y captación, con un enfoque de innovación que le permite ofrecer servicios bancarios a 19.1 millones de clientes.

Es una empresa controladora filial del Banco Español, uno de los grupos financieros líderes en Europa y considerado entre uno de los más grandes de la Zona Euro. El Banco Español es un grupo financiero con una elevada solvencia y rentabilidad, tiene presencia en 32 países del mundo, destacando su compromiso con la región latinoamericana donde forma la franquicia financiera líder.

1.2 Estructura del modelo de negocios.

El modelo de negocios del Grupo Financiero Bancario consiste en la atención y distribución especializada de productos y servicios para cada tipo de cliente. El Grupo Financiero mantiene una filosofía de control de riesgo y tiene un objetivo de crecimiento rentable en el largo plazo.

El Grupo trabaja por un futuro mejor para las personas, al ofrecer a su clientela una relación de beneficio mutuo, servicio proactivo, asesoramiento y soluciones integrales tanto para individuos o empresas.

También participa de manera relevante en el mercado de servicios financieros no bancarios con negocios líderes en los segmentos de banca seguros, gestión de activos, gestión de fondos para el retiro, transferencias de remesas de dinero y fondos de inversión.

1.3 Historia.

1932: Se funda la Empresa en la ciudad de México bajo el nombre de Banco de Comercio.

1982: El gobierno mexicano nacionalizó la banca mexicana, incluyendo a la Empresa.

Diciembre 1991: Con la privatización bancaria, se forma Grupo Financiero Empresarial que retoma el control de la Empresa.

Junio 1995: Se crea el negocio de transferencias de dinero, bajo el nombre de La Empresa Transfer Services.

Diciembre 1996: Se crea la administradora de fondos de pensiones del Sistema de Ahorro para el Retiro bajo el nombre de Afore Empresa, así como la empresa especializada en banca seguros llamada Seguros Empresa.

Mayo 1997: Se crea el negocio de pensiones bajo el nombre de Pensiones Empresa.

Julio 2000: Banco Español capitaliza a La Empresa con USD 1,400 millones y nace Grupo Financiero Español Empresa (GFEE) de la fusión de Grupo Financiero Español Probursa con Grupo Financiero Empresa.

Agosto 2000: GFEE adquiere Banca Promex, un banco con fuerte presencia en la región centro y occidente del país y se consolida la mayor institución financiera de México.

Junio 2002: Grupo Financiero Español Empresa se convierte en sociedad controladora filial de Banco Español como consecuencia de que este último llegó a detentar más del 51% de las acciones representativas del capital social del Grupo.

Febrero – junio 2004: Español lanza una Oferta Pública de Recompra por aproximadamente la tenencia accionaria restante de Español Empresa en manos de inversionistas minoritarios, logrando una tenencia total al cierre de la operación de 98.88%.

Enero 2005: Grupo Financiero Español Empresa adquiere a Hipotecaria Nacional, importante sofo hipotecaria que al unirla con la red hipotecaria de La Empresa se creó la Banca Hipotecaria, misma que quedó conformada por 120 sucursales especializadas.

Julio 2009: Se lleva a cabo la fusión de Español Empresa como fusionante y Español Empresa Servicios como fusionada, con la autorización de la Comisión Nacional Bancaria y de Valores, así con la aprobación del Banco de México.

Noviembre 2012: El 27 noviembre de 2012, Banco Español, S.A. acordó con Afore XXI Banorte, S.A. de C.V. el plan de venta de acciones de Afore Español Empresa, S.A. de C.V. El cierre de la operación quedó sujeto a la obtención de la autorización tanto de la Comisión Nacional de Ahorro para el Retiro (CONSAR) como de la Comisión Federal de Competencia (COFECO), mismas que fueron otorgadas en noviembre y diciembre de 2012 respectivamente. Con ello, la venta se formalizó el 9 de enero de 2013 y en esa misma fecha fue trasladado la Administración y control de la Sociedad.

Enero 2013: el 9 de enero de 2013 se concretó la venta de la subsidiaria Afore Español Empresa a Afore XXI Banorte, trasladando en esta fecha la administración y control de la sociedad.

Febrero 2013: El 22 de febrero de 2013, el Grupo Financiero Español Empresa realizó la compra del 20.61% de la tenencia accionaria de Seguros Español Empresa a Banco Español. Dicha operación se realizó a efecto de que el GFEE consolide su participación accionaria en Seguros Español Empresa para alcanzar una tenencia de 99.99%.

Junio 2016: Venta de las subsidiarias de La Empresa USA.

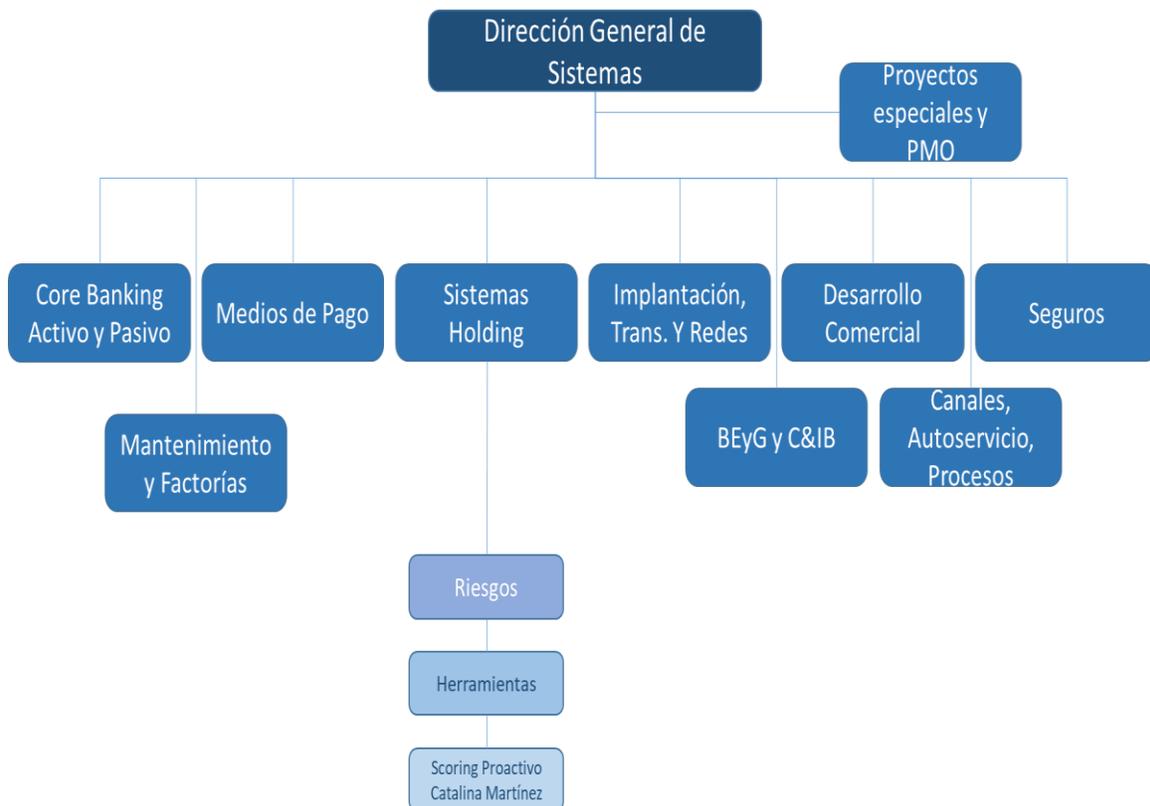
Febrero 2017: El 14 de febrero de 2017 se llevó a cabo la fusión de Hipotecaria Nacional, BETESE y Desitel con el Banco.

1.4 Organigrama.

El organigrama de una empresa es muy importante para asegurar la comunicación interna y el flujo de trabajo, así como dar a conocer la estructura jerárquica. Desde el presidente, pasando por los directores, gerentes y empleados.

En la figura 1.1 se presenta en organigrama de la dirección general de sistemas, enfocando al área de Sistemas Holding, dirección a la cual pertencí durante el desarrollo de los proyectos que se mencionarán en este reporte:

Figura 1.1 Organigrama



Capítulo 2 –Metodología y Descripción de Proyectos.

En mi trayectoria profesional dentro de la empresa me desarrollé profesionalmente en los siguientes puestos:

- Becario.
- Ingeniero Entrenante.
- Analista de Sistemas.
- Líder de Proyecto.
- Gerente de Proyectos.
- Consultor Funcional y Técnico.

Dentro de los perfiles anteriores las actividades principales fueron documentar, programar, diseñar, optimizar, instalar, dirigir, supervisar y coordinar proyectos para el diseño y desarrollo de los sistemas.

Con el tiempo fui incrementado mis conocimientos en los aspectos informáticos, del negocio bancario, así como mis habilidades para su desarrollo de software, acrecentando mi capacidad de análisis de situaciones reales y planteamiento de soluciones a los problemas detectados sobre el manejo de los datos dentro de la organización.

Los proyectos elegidos a describir en el presente reporte fueron los más relevantes en orden de aparición en mi vida profesional:

- Modelo de Datos para Campañas de Telemarketing
- Cédula Comparativa del Empleado
- Scoring Proactivo Global.
- Modelo Detección de empresas con potencial de crecimiento.

Cabe mencionar que proyecto Modelo de datos para campañas de telemarketing fue mi primer reto dentro de la empresa.

2.1 Metodología Corporativa.

A continuación, se describe brevemente la Metodología Corporativa que fue el marco de referencia utilizado para los desarrollos de software dentro de la empresa, esta metodología me permitió estructurar, planificar y controlar los proyectos, y así generar desarrollos ordenados, claros y específicos, con el tiempo fue un apoyo para ampliar mis conocimientos, habilidades y paciencia en el desarrollo de proyectos.

La Metodología Corporativa buscaba proporcionar una comunicación clara y oportuna entre las distintas áreas que participaban en el ciclo de vida productivo del

desarrollo. La metodología estaba basada en **Capability Maturity Model Integration** (CMMI), la cual se adaptó e implanto en la institución. La empresa realizó una inversión en la capacitación del personal de las diferentes áreas de sistemas enfocadas al desarrollo de software, que con el tiempo fue madurando, y se fueron obteniendo desarrollos de software con controles y mejor planeación.

El objetivo del desarrollo de software era transformar las necesidades del cliente en un producto de solución informático que fuera utilizado por los usuarios a quienes iba dirigido y que cumplió con las siguientes expectativas:

- Entrega con la calidad pactada.
- Entrega en el plazo establecido.
- Realizado con el menor costo posible.

La metodología corporativa estaba dividida en tres grandes conjuntos de procedimientos:

- Gestión de iniciativas.
- Desarrollo.
- Gestión del servicio.

La figura 2.1 se muestra el flujo general de la metodología corporativa y la relación de las fases:

Figura 2.1 Diagrama General

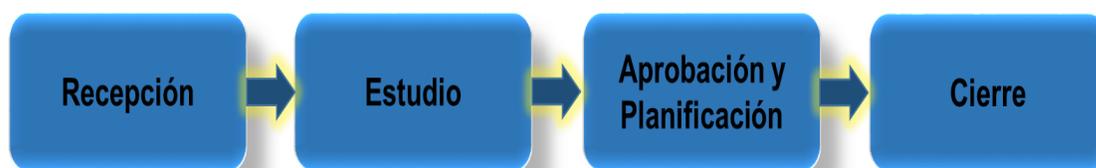


En seguida describiré brevemente en que consiste cada conjunto de procedimientos.

2.1.1 Gestión de Iniciativas.

Este conjunto de procedimientos consistía en recoger las necesidades de las áreas de negocio, definiendo su viabilidad o inviabilidad con base en las posibles opciones de desarrollo, costos e infraestructura necesaria para su implementación. La gestión de iniciativas se desarrollaba a nivel dirección, en la figura 2.2 se muestra el flujo de actividades que se realizaban en esta sección:

Figura 2.2 Flujo de Gestión de Iniciativas



Recepción:

Recoger las necesidades de las Áreas de Negocio, incorporándolas a la cartera de pedidos e identificando a todos los grupos implicados en su solución. Identificar los objetivos a cubrir en relación con la iniciativa recibida.

Estudio:

Estudiar y analizar la necesidad recogida, para valorar el esfuerzo de ejecución y acometer su realización.

Aprobación y Planificación:

Aceptar o denegar la iniciativa en función de la valoración de la iniciativa y de la disponibilidad de presupuesto, así como reclasificarlo como proyecto si se considera necesario.

Realizar la priorización y planificación de la cartera de iniciativas como resultado del análisis de urgencia, fechas de obligado cumplimiento, posibles concurrencias e importancia global para la organización.

Cierre:

Es realizar el cierre formal de la iniciativa una vez se haya realizado la misma.

2.1.2 Desarrollo.

En el desarrollo se encuentran las fases propias llevadas a cabo para desarrollar y poner en marcha un Sistema de Información, las actividades de esta fase son sistemas completos de técnicas que incluyen procedimientos paso a paso, productos resultantes, funciones, herramientas y normas de calidad para la terminación del ciclo de vida completo del desarrollo de sistemas. En la figura 2.3 se describen las actividades núcleo del diseño y desarrollo de software:

Figura 2.3 Fases de Desarrollo



Análisis:

Delimitar el alcance del sistema, mediante modelos iniciales de alto nivel y detallando el aspecto que tendrá el nuevo sistema de cara al usuario, así como identificar los requisitos de infraestructura para implementar la solución deseada.

Diseño:

Obtener un diseño detallado de los componentes del sistema y de la Interfaz de Usuario a partir de la documentación generada durante la fase de Análisis.

Construcción:

Construir los elementos de software que forman el sistema y que se han diseñado en las fases anteriores.

Preparar los entornos para cada una de las pruebas y realizar las pruebas unitarias correspondientes de cada componente del sistema.

Pruebas funcionales:

Realizar los distintos tipos de pruebas funcionales, de sistema e integración, verificando que la aplicación cumple con los niveles de calidad requeridos

Pruebas de usuario:

Realizar las pruebas de aceptación validando que satisface los requisitos establecidos por el usuario y preparar los entornos de formación y producción

Implantación

Realizar la implantación de los programas, módulos, componentes y demás elementos que forman el sistema, y que han sido diseñados, contruidos, probados y validados en las fases anteriores.

2.1.3 Gestión del Servicio

El objetivo de Gestión de Servicios es asegurar que los procesos adecuados, las personas y la tecnología están en su lugar para que la empresa pueda cumplir con sus objetivos de negocio. Una buena Gestión de Servicios consiste en realizar todas aquellas actividades y tareas que permitan:

- La puesta en marcha del proyecto.
- Seguimiento y Control.
- Comprobación del correcto desempeño.
- Cierre y finalización del proyecto.

La figura 2.4 muestra el flujo de las fases generales involucradas en la gestión de Servicios:

Figura 2.4 Flujo de Gestión del Servicio



Inicio:

Cada año la empresa destina el presupuesto acorde a los proyectos autorizados en su plan anual, las áreas de infraestructura involucradas generan sus planes e inversiones en recursos de tecnología de información para el apoyo a los procesos de negocio.

Seguimiento y Control:

Los productos y servicios están orientados a la implantación de sistemas alineados con los objetivos del negocio, de punta a punta desde las fases diagnóstica y de planificación hasta la implantación, monitoreo y supervisión. Las áreas involucradas de Infraestructura dan seguimiento y control a sus planes de adquisición de insumos. Determinan todos los procesos relativos a la producción, administración y uso de servicios de Tecnologías de Información (TI).

Cierre:

Una vez realizada la instalación de Sistema de Información y transcurrido el periodo de garantía, se procede al cierre de la gestión del proyecto.

Soporte:

El servicio de soporte técnico sirve para ayudar a resolver los problemas que puedan presentarse a los usuarios, mientras hacen uso de los sistemas de información.

A continuación, explicaré el primer proyecto que describiré en este reporte.

2.2 Proyecto 1 – Generación Modelo de Datos.

Este primer proyecto consistió en la generación de un modelo de datos dentro de Customer Knowledge DataBase (CKDB), participé en este proyecto siendo analista de sistemas, las actividades que desarrollaba hasta este momento consistían en generación de reportes solicitados por los usuarios del área de Inteligencia Comercial.

Mi líder de proyecto me seleccionó para participar en este proyecto, permitiéndome diseñar el modelo de datos, generar el plan de trabajo e implantar el diseño y desarrollo de los programas en la Base de Datos.

La duración del proyecto fue aproximadamente de ocho meses, incluyendo el periodo de garantía del software implantado.

A continuación, describo a mayor detalle el proyecto.

2.2.1 Antecedentes.

La Empresa como líder en el mercado ha enfocado sus esfuerzos en mantener la atención en sus clientes ofreciéndole productos y servicios que los mantengan satisfechos, una de las problemáticas a las que se ha enfrentado es el contacto directo e instantáneo con ellos, para lograr esto ha diseñado la estrategia de campañas dirigidas de telemarketing, las cuales permiten un contacto bidireccional entre el banco y el cliente, ahorrando costos, incrementando la colocación de sus productos financieros, abarcando un mercado más grande y con mayor diversidad de productos, extendiendo la fuerza de ventas con costos mínimos y mejorando el servicio a los clientes.

2.2.2 Objetivo

Crear un modelo de datos para las campañas de telemarketing que permita el control de las campañas generadas, los clientes a las que van dirigidas, las respuestas recibidas de los mismos, así mismo generar reportes estadísticos que ayuden al negocio a la toma de decisiones para generar mejores estrategias.

2.2.3 Descripción

Este proyecto se enfocó en generar un modelo de datos basado en las diferentes campañas de telemarketing de la banca comercial, para almacenar el histórico de las campañas, así como de las respuestas recibidas por los clientes, lo anterior con el fin de tener un fácil manejo de la información, un almacenamiento masivo de datos y un reporte más ágil y certero del comportamiento de las campañas y su impacto entre los clientes.

Anteriormente los archivos de campañas eran generados en un archivo plano y entregadas por el usuario de negocio a la empresa de call center (cumpliendo las políticas de transferencia de información a empresas externas) para la aplicación de las campañas entre los clientes, la empresa de call center realizaba la campaña de acuerdo a las fechas establecidas y entregaba de la misma forma las respuestas recibidas por el cliente. El control de la información recaía en la administración del usuario de negocio, como las campañas eran periódicas en ocasiones no terminaba de procesar la información recibida de unas campañas cuando llegaban las siguientes, generando un cuello de botella en la entrega de reportes para la toma de decisiones por parte del negocio, así como molestia y reclamos por parte de los clientes que eran contactados en más de una campaña al mismo tiempo, entre otras cuestiones.

En la figura 2.5 se muestra el flujo general de campañas.

Figura 2.5 Flujo de Campañas de Telemarketing



2.2.4 Solución

Como punto de partida se realizaron las reuniones con los usuarios involucrados para escuchar sus necesidades, estas actividades formaron parte de la gestión de iniciativas de la metodología corporativa, específicamente el área de gestión del servicio.

Se levantó un requerimiento (formato A825 – Formalización de la iniciativa), identificando la necesidad del negocio, los tiempos deseables de atención y los usuarios involucrados en la petición.

El socio de negocio (pertenece a la Gestión del Servicio) identificó al área de sistemas que atenderá la petición (dirección específica de sistemas) y entregó la petición de manera oficial con los requerimientos solicitados por el usuario (formato C100 – Solicitud de proyecto), debidamente firmado por los usuarios involucrados en la petición

El área de sistemas CKDB recibió la petición y la canalizó al equipo de trabajo destinado para la solución, el líder de proyecto convocó a reuniones con el usuario para realizar mesas de trabajo y poder entender el proceso actual e identificar la posible solución, se diagramó el flujo del proceso de campañas, se identificaron las fuentes de información, se ubicaron los equipos donde se almacenaba, así como los procedimientos que realizaba el usuario para realizar la gestión de campañas, se solicitó una muestra de la información para su análisis y revisión, mi trabajo dentro de estas reuniones fue generar las convocatorias, generar las minutas de las reuniones con los avances y pendientes, generar el flujo del proceso de campañas (figura 2.5), así como la revisión y análisis de la muestra de información solicitada al usuario.

Una vez que se analizaron los detalles requeridos en la solicitud, se plasmó la solución a la necesidad de negocio en el formato C102 (Solución de proyecto), especificando de manera clara y concisa el alcance, así como las fechas de atención y el costo que implicaría este desarrollo, una vez cerradas las dudas y estando los 3 frentes de acuerdo se pasó a firmas, una vez firmado por todos los involucrados se notificó a la oficina de proyectos, con lo cual se formalizó el inicio del proyecto (Desarrollo), mi trabajo consistió en elaborar este formato.

Una vez aprobado el desarrollo, generé el plan de trabajo detallando las actividades de análisis, diseño, construcción, pruebas funcionales, pruebas de usuario e implantación a desarrollar (archivo Microsoft Project) con tiempo y costo. Se dio de alta ante la oficina de proyectos para su seguimiento semanal (Gestión de Servicios), la oficina de proyectos era la entidad encargada de verificar el avance en tiempo y forma de cada una de las actividades del plan de trabajo, esta área generó el registro de avances, retrasos y alertas de manera semanal y envió de reportes a dirección general de sistemas.

Las actividades en que participé adicionales al desarrollo fueron la gestión de espacios, tablas, permisos y accesos, el usuario participó en las pruebas de usuario y proporcionó su visto bueno para la implantación del desarrollo en el ambiente productivo previamente gestionado.

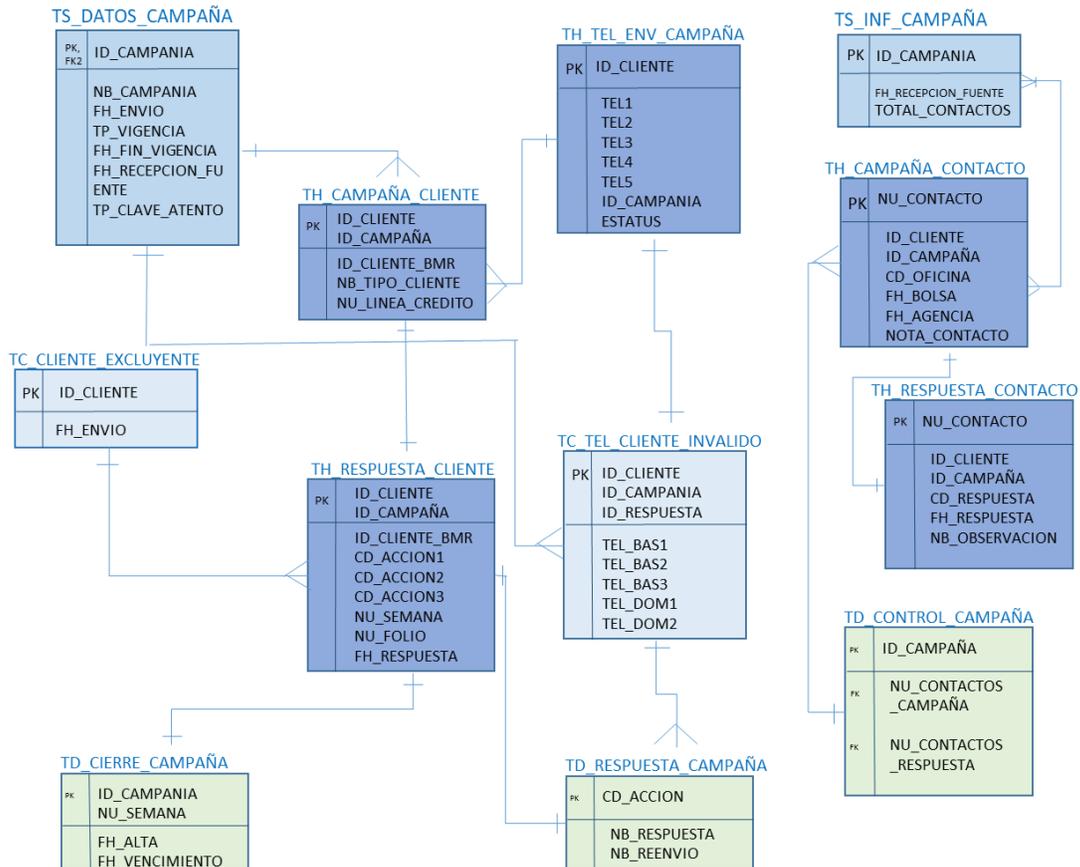
Una vez que se obtuvo el Visto Bueno o Aprobación de la implantación se gestionó el cierre el proyecto ante la oficina de proyectos, está gestión implica contar con el formato correspondiente firmado por parte del usuario y el acta de implantación del área de producción debidamente llenada y firmada, una vez concluido el periodo de garantía de soporte.

La figura 2.6 muestra el flujo implementado como solución de este proyecto.



El modelo de datos que diseñé se muestra en la figura 2.7.

Figura 2.7 Modelo de Datos campañas TMK



El modelo de campañas consistió en 4 tipos de tablas:

- TH – tablas de hechos.
- TC – tablas a nivel cliente.
- TD - tablas de dimensiones o catálogos.
- TS – tablas resumizadas.

Las tablas principales del modelo fueron TH_CAMPAÑA_CLIENTE era la tabla de hechos que almacenaba cada campaña desarrollada, la tabla TH_RESPUESTA_CLIENTE almacenaba las respuestas recibidas de los clientes, la tabla TH_TEL_ENV_CAMPAÑA almacenaba los teléfonos de los clientes generados en la campaña, la tabla resumizada TS_DATOS_CAMPAÑA donde se almacenaba la información general de la campaña y los catálogos TD_RESPUESTA_CAMPAÑA tabla donde se almacenaban las respuestas posibles del cliente y TD_CIERRE_CAMPAÑA donde se almacenaba la información de alta y vencimiento de la campaña.

Es importante mencionar que este modelo se desarrolló bajo una plataforma Oracle 7 y en un ambiente semi-productivo, automatice las interfaces de entrada y salida mediante store procedure, triggers y shell's, así como los reportes periódicos solicitados, es decir, una vez recibidas las reglas de negocio a aplicar para la campaña, estas se parametrizaban en el proceso y se calendarizaba su ejecución, una vez generado el archivo de la campaña, era depositado en un servidor Windows, donde el área de transmisiones externas lo tomaba y realizaba su encriptación y respectivo envío al empresa que ejecutaba las campañas, dado que las campañas tenían duraciones diferentes, todos los días un proceso automático revisaba el servidor Windows en busca de posibles respuestas de campañas, en caso de encontrar, se disparaba un proceso que cargaba dicha información de manera automática dentro del modelo, para mantener actualizada la información desarrollé triggers que garantizaban la no duplicidad de datos, evitando de esta manera la manualidad, dado que en ocasiones los archivos de respuestas estaban incompletos y era necesario volver a realizar su carga.

Así también se generaban reportes automáticos de manera mensual, estos reportes se encontraban definidos, una vez generados eran depositados en el servidor Windows para dejarlos a disposición de los usuarios, en caso de que se requiriera un reporte diferente, éste se realizaba de manera manual.

2.2.5 Resultados

El modelo de datos permitió un fácil y rápido manejo de los datos, dado que la información se encontraba almacenada en un solo lugar, evitando la dependencia manual, dentro de un ambiente productivo que contaba con niveles de servicio,

permitiendo mejorar los reportes entregados de manera periódica, así como como los reportes ad-hoc solicitados por el área usuaria. Permitiendo que el conocimiento se distribuyera y no quedara bajo el resguardo de un par de personas.

Una vez implementada la solución, el área de negocio de inteligencia comercial visualizo la necesidad de generar las campañas de manera automática de acuerdo a una lista de parámetros que ellos podrían modificar de acuerdo a las características del mercado, esta segunda petición dio como resultado otro proyecto, así también se solicitó el modelo de datos de las campañas de Marketing dado el éxito de este modelo.

Este proyecto fue catalogado de prioridad baja, por lo cual las personas involucradas del área de sistemas Holding fueron dos: el líder de proyecto y un analista, esta última persona fui yo.

2.3 Proyecto 2 – Cédula Comparativa.

A continuación, explicaré el proyecto Cédula comparativa del empleado, mi rol en este proyecto fue como líder de proyecto. Este proyecto consistió en presentar una ficha de información que mostrara las circunstancias personales/profesionales de cada empleado a las áreas locales de Recursos Humanos, en un formato sencillo y fácil de utilizar por todas las áreas.

Como líder de proyecto seleccione a dos colaboradores para el diseño y desarrollo, uno con perfil de BD y otro con perfil de SAS, el primer colaborador desarrollaría las consultas para la extracción de información de la Base de Datos de Recursos Humanos y el segundo se enfocaría en el diseño y desarrollo de la ficha y su visualización en Excel, mediante el complemento SAS para Microsoft Office, ya que una de los requerimientos solicitados en este proyecto fue utilizar un software que ya existiera en la empresa y de fácil uso para cualquier empleado de RH.

En este proyecto se me permitió por primera vez organizar y gestionar un proyecto con un equipo de trabajo a mi cargo, controlar el presupuesto, riesgos y actividades de mis colaboradores.

La duración del proyecto fue aproximadamente de 10 meses, incluyendo el periodo de garantía del software implantado.

2.3.1 Antecedentes.

La Empresa siendo un grupo de servicios financieros bancarios y no bancarios posee una gran plantilla de empleados en sus diferentes empresas, el área de recursos humanos se enfoca a que los esfuerzos estén alineados al negocio, por ello, busca la creación de valor sostenible, asegurando cohesión y consistencia hacia los mismos objetivos del negocio y estableciendo planes de acción específicos para mejorar los niveles de servicio y resultados del Capital Humano. Fomentando y promoviendo el cumplimiento de las líneas estratégicas de la Dirección de Recursos Humanos con los empleados del Grupo con el objetivo de mejorar el clima laboral a través de incrementar la Satisfacción, Motivación y Compromiso de los Empleados, para lograr lo anterior se enfoca en los siguientes puntos:

- Cercanía: Recursos Humanos cerca de cada empleado y cada Unidad.
- Comunicación: Llegamos a todos los empleados y abiertos a escucharlos.
- Transparencia: Hacer pública la información y las prácticas de Recursos Humanos.
- Eficiencia: Maximizar la capacidad del equipo de RR HH.
- Servicio: Establecer una Cultura de Servicio que genere valor al negocio.
- Compromiso: Convertirnos en el Mejor lugar para Trabajar.

Los cimientos del Grupo están en su Cultura Corporativa, así como el valor y la reputación de su marca.

2.3.2 Objetivo

Proveer al área de Gestión RR HH de Información útil de candidatos para cobertura de Puestos de una manera ágil y oportuna, generando una ficha comparativa del empleado con todos los datos de su trayectoria dentro de la empresa, esta ficha deberá contar con un esquema de seguridad para garantizar la confidencialidad de la información mostrada, para que los gestores de RRHH puedan evaluar y dar seguimiento a cada uno de los empleados asignados.

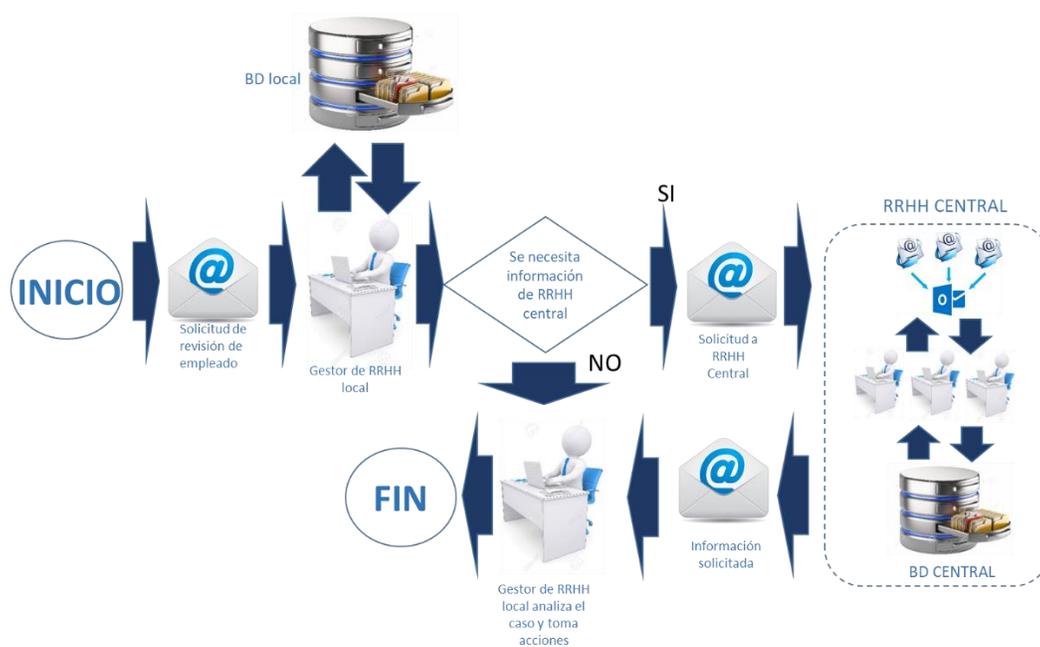
2.3.3 Descripción

En este proyecto se solicitó generar un reporte con los datos del empleado, el cual debería contener sus datos personales y profesionales para que el gestor de recursos humanos asignado contará con la información necesaria y dar seguimiento puntual al desarrollo del empleado, generando líneas de acción de acuerdo a las estrategias del Banco.

Anteriormente alguna área local solicitaba la revisión de empleados específicos al gestor local de recursos humanos, éste realizaba la extracción de la información de la base de datos local, para realizar la revisión solicitada, en caso de que tuviera la información necesaria en el procedimiento a aplicar, el gestor de recursos humanos generaba las acciones pertinentes. Sin embargo, existían ocasiones que la información no era suficiente para el procedimiento, en ese caso se enviaba la solicitud de información al área central, este tipo de peticiones llegaban de todas regiones a nivel nacional, por lo cual el área central se veía sobrepasada en su capacidad instalada, tardaban días en extraer la información y en el envío de su respuesta. Para tratar de reducir estos tiempos las áreas locales generaban archivos que guardaban de manera local en sus equipos, por lo que el riesgo de tener datos no actualizados era muy grande, lo anterior provocó acciones incorrectas de parte de los gestores de recursos humanos locales.

En la figura 2.8 se muestra el flujo para la obtención de la cédula del empleado antes del proyecto.

Figura 2.8 Obtención Cédula Comparativa del empleado



El usuario de recursos humanos central solicitó que la herramienta de visualización se debería considerar lo siguiente:

- La extracción de la información debería realizarse en línea.
- Fácil uso para los gestores.
- Sin instalación de SW.
- Con sistema de autenticación para entrar al sistema.
- Con tiempo de respuesta aceptable.

2.3.4 Solución

Como punto de partida se realizaron las reuniones con los usuarios de recursos humanos para escuchar sus necesidades, estas actividades formaron parte de la gestión de iniciativas de la metodología corporativa, específicamente el área de gestión del servicio.

El área de gestión del servicio mediante su Socio de Negocio o BP (Business Partner) levantó un requerimiento (formato A825 – Formalización de la iniciativa), identificando la necesidad del negocio, los tiempos deseables de atención y los usuarios involucrados en la petición. Se realizaron varias reuniones para cerrar el alcance del requerimiento, lo anterior se plasmó en el formato C100 (formato de solicitud de proyecto) y se firmó con el Visto Bueno de los implicados.

El BP entregó el C100 al área de sistemas Analítica Central, quien tuvo 10 días de acuerdo a la metodología corporativa para entrar el diseño de su solución, la cual plasmó en el formato C102 (Solución de Proyecto), para el cierre de este formato convoque a varias reuniones con el usuario de recursos humanos para cerrar los detalles específicos de la solución propuesta, incluyendo el plan de trabajo del Desarrollo del proyecto, en estas reuniones asistí con mi equipo de trabajo para la aclaración de dudas técnicas del usuario, uno de los puntos importantes para el cierre de la solución fue acotar la implantación de la solución a solo 5 oficinas locales de recursos humanos pertenecientes a la región centro del país, debido a que el área de CCR (Centro corporativo Regional) sería la encargada de esta tarea, era necesario un plan que incluyera dicha instalación masiva a nivel nacional, por lo que esta distribución se realizó en una segunda fase del proyecto, donde solo fungimos como apoyo en la capacitación de los usuarios y posibles dudas en la instalación de software. Una vez revisado y aceptado el alcance, se procedió a la firma del formato, iniciando formalmente el desarrollo del proyecto.

Dentro del análisis realizado en el desarrollo, uno de los puntos cruciales de este proyecto fue determinar que herramienta era la que se utilizaría, debido a que era necesario evitar la compra de un nuevo software (ya que no se contaba con presupuesto de infraestructura dentro de este proyecto), adicional a que esto retrasaría considerablemente el desarrollo al tener que realizar las gestiones correspondientes con el área de CCR (Centro Corporativo Regional), por lo que nos enfocamos a buscar alguna herramienta ya instalada en la empresa en un ambiente productivo que pudiera solventar la necesidad del usuario; Nos encontramos con el Web Report Studio (WRS) de SAS, que estaba dentro de los requisitos que el usuario solicitada así como dentro del presupuesto y tiempos esperados. No fue posible evitar la instalación de software en los equipos de los gestores locales, sin embargo, esta instalación era mínima y compatible con Microsoft Office.

Dentro del Web Report Studio se encuentra un elemento llamado Stored Process que podía fungir como transacción en línea, es decir que cuando los diferentes gestores necesitaran extraer información del BD central, el Stored Process iría en ese momento por la información a la Base de datos y la desplegaría en pantalla, una vez generado el reporte este podía copiarse a Excel que es la herramienta de uso común que utilizan los gestores para sus análisis y seguimientos.

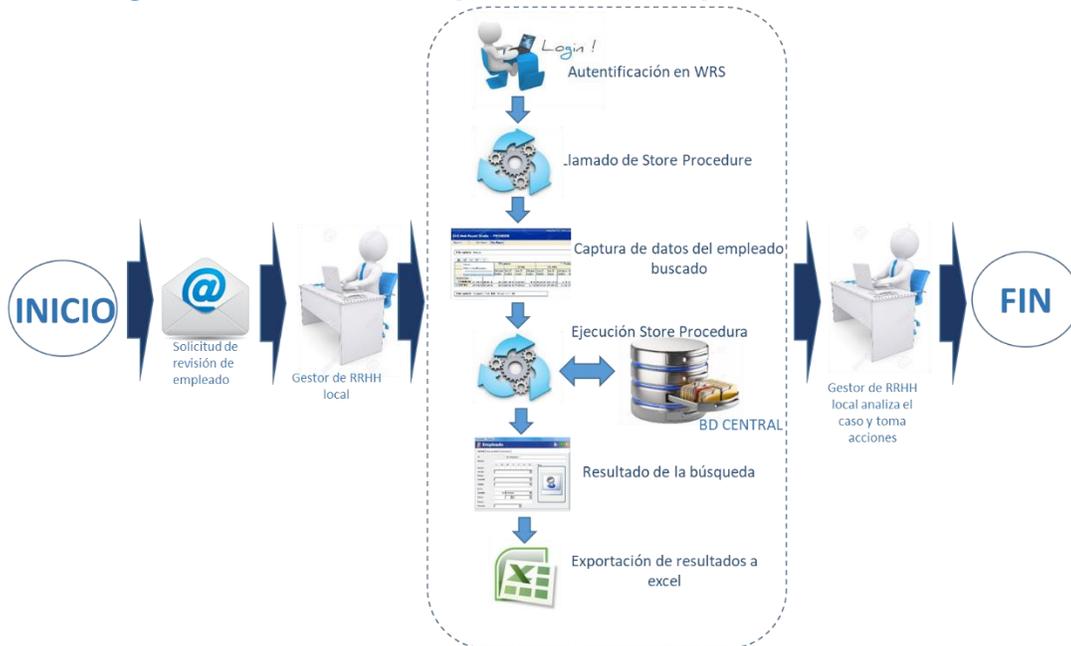
Otra actividad que realicé en el análisis fue revisar el tema del licenciamiento de esta aplicación, se determinó que no era necesario la compra de nuevas licencias, enfocándonos entonces solo en la instalación del SW, al ser una aplicación WEB solo era necesario abrir los puertos específicos para que los gestores del interior de la republica pudieran acceder al servidor central, este tema los revisé con el área de CCR en algunas reuniones de trabajo, quien nos indicó el esquema y las gestiones correspondientes para este fin, así como también la calendarización de las instalaciones.

Una vez cerrados los temas de infraestructura mi equipo y yo nos enfocamos en la fase de diseño, mantuvimos varias sesiones de trabajo donde mis colaboradores me presentaron sus propuestas de solución, revisamos la solución al reporte, se analizaron las fuentes de información y se determinó el flujo del proceso. Cada analista se enfocó a las actividades de su herramienta y la comunicación entre ambas.

El diseño de la solución generado consistió en que el gestor de recursos humanos local recibiría las solicitudes de revisión de empleados vía correo electrónico como hasta el momento se venía haciendo, una vez que el gestor identificara a el empleado o empleados a revisar, abriría Microsoft Excel, utilizando SAS Add-In for Microsoft Office se firmaría en la aplicación WRS, una vez conectado a la aplicación llamaría al Store Procedure desarrollado para extraer el reporte, una vez ejecutado el Store Procedure, éste mostraría una ventana para la captura del número de empleado o empleados a buscar, el procedimiento ejecutaría la consulta a la base de datos central y mostraría los resultados en la pantalla del gestor, el reporte mostrado podía ser exportado a Excel por parte del usuario si así lo decidiera y continuaría con su análisis del empleado y aplicaría las acciones correspondientes, Esta ejecución tardaría unos minutos y no días como anteriormente sucedía.

En la figura 2.9 se muestra el flujo de la solución diseñada en este desarrollo.

Figura 2.9 Cédula Comparativa del Empleado



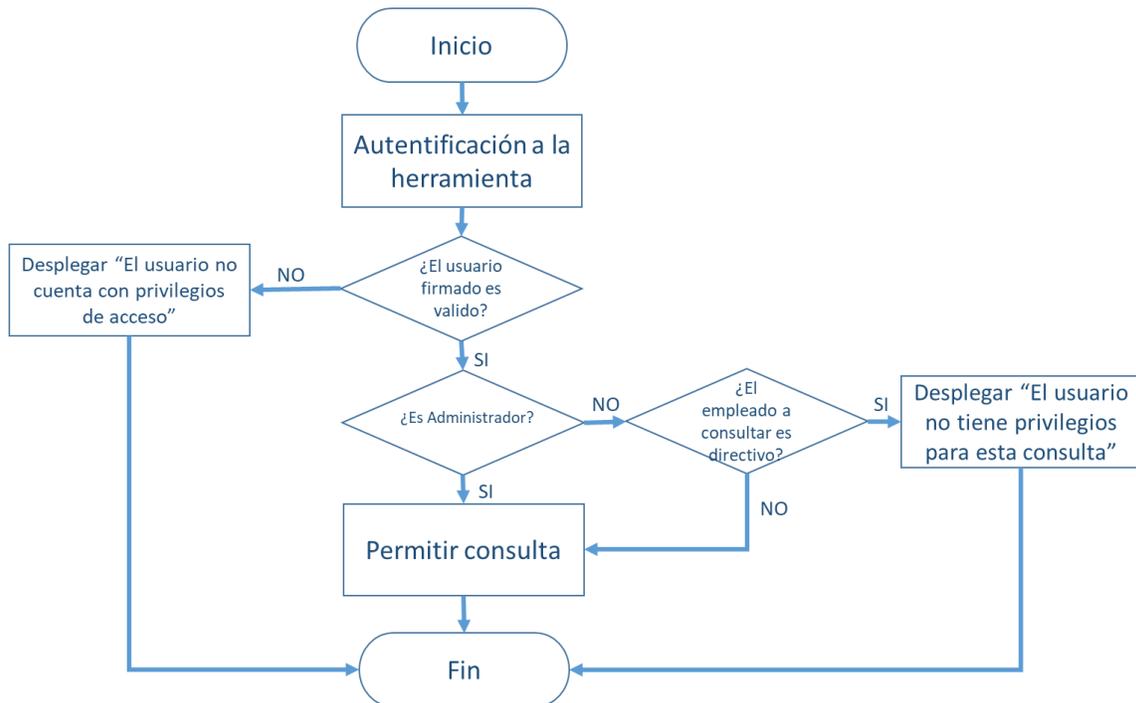
Es importante mencionar que la información utilizada en el desarrollo era considerada sensible, por lo anterior RH solicito se firmara un acuerdo de confidencialidad adicional (cada empleado firma un acuerdo de confidencialidad sobre el manejo de la información al entrar a trabajar en la empresa), este acuerdo

lo firmamos todos los participantes donde el compromiso fue no realizar ningún mal uso de la información y las consultas realizadas solo serían sobre la lista de empleados proporcionados por el área de RH (en pruebas funcionales), la lista enviada por RH fue de solo 10 empleados, adicional que los usuarios utilizados para las consultas en el desarrollo fueron monitoreados de manera puntual por el área de seguridad de la información.

Durante el Diseño también se automatizó el esquema de seguridad para las consultas establecido por el área usuaria, dicha automatización fue diseñada por mí, la autenticación de los gestores se realizaría mediante el uso del Store Procedure de la cédula comparativa, en función de los perfiles de usuario que se tenían establecidos en el área de recursos humanos.

En la figura 2.10 se muestra el esquema de seguridad diseñado, una vez firmado el gestor de recursos humanos a la herramienta se validaría si el usuario utilizado es válido, en caso negativo se desplegaría un mensaje de “El usuario no cuenta con privilegios de acceso”, en caso contrario se realizaría la siguiente validación, el usuario es administrador, en caso afirmativo las consultas no tendrían restricciones, en caso negativo se realizaría una validación más sobre los empleados a consultar, en caso de empleados directivos solo el administrador podía realizar la consulta, por lo que si un gestor no contaba con privilegios de administrador aparecería el mensaje “El usuario no tiene privilegios para esta consulta”. La información sobre los privilegios de cada gestor de recursos humanos se encontraba cargada en una tabla de la base de datos.

Figura 2.10 Esquema de Seguridad para Gestores de RH



Anterior a este proyecto, se desarrolló el modelo de datos de RH denominado LARH (Laboratorio Analítico de Recursos Humanos), donde se concentró toda la información de los empleados del Grupo, debido a que se encontraban en diferentes repositorios dificultando su administración y el manejo de la información, este modelo sirvió como fuente de información para el proyecto.

En la fase de construcción del Desarrollo, mi participación fue dar seguimiento a la creación de los componentes que mis colaboradores desarrollaron en las plataformas SAS y Oracle de acuerdo a plan de trabajo, en caso de dudas o riesgos encontrados, manteníamos reuniones de trabajo para encontrar la solución y continuar con el avance del proyecto.

En las pruebas funcionales mis actividades fueron las de generar los casos de uso, casos de prueba y realizar la validación de los componentes desarrollados, dando mi aprobación para pasar a las pruebas de usuario.

En las pruebas de usuario mis actividades fueron realizar reuniones de trabajo con los usuarios involucrados en las pruebas, de manera presencial para los gestores centrales y vía telefónica para los gestores locales de la región centro, se generó un plan de pruebas de usuario, al cual le dí seguimiento puntual a su ejecución para cumplir con el plan del proyecto en tiempo y forma.

Una vez obtenido el VoBo de parte del usuario en las pruebas de usuario, se llevó a cabo la implantación de la aplicación en Producción, estas actividades fueron gestiones con el área de CCR (producción distribuida), para que los componentes desarrollados pasaran a las áreas productivas respectivas (Base de datos y servidor SAS), estas gestiones fueron realizadas por mi equipo de trabajo bajo mi supervisión, una vez realizado lo anterior, el usuario utilizó la aplicación instalada bajo nuestra supervisión (mía y de mi equipo) para la atención a cualquier incidente presentado. Para este proyecto la única incidencia fue la desconexión de la aplicación en 2 oficinas de las 5 que se incluyeron en la implantación, pero los problemas se deberían a los canales de comunicación (aspecto perteneciente al CCR) y no a la aplicación desarrollada. Por lo cual se firmó el acta de liberación del proyecto, dando paso al área de Gestión del Servicio para la atención y mantenimiento de la aplicación en lo subsecuente.

2.3.5 Resultados

Los gestores contaron con la información actualizada de los Empleados que se consideren como candidatos para la cobertura de algún puesto en particular, pudiendo realizar un comparativo rápido de los datos y trayectoria de los mismos, con la facilidad de ingresar uno o más registros para obtener el comparativo entre los homólogos o candidatos al puesto.

La distribución de la cedula comparativa se realizó por fases por las diferentes infraestructuras manejadas en las diferentes oficinas, lo anterior por recomendación del área de CCR, esta distribución duro alrededor de 8 meses.

La primera oficina que utilizo la herramienta fue el área central de RH, quien a pesar de seguir recibiendo las solicitudes de las oficinas del interior el proceso le ayudo a disminuir los tiempos de atención, aproximadamente en un 80%.

En la Figura 2.11 se muestra el diseño de la Cédula Comparativa.

Figura 2.11 Cédula Corporativa

0_CAMPO	MEX00001	MEX0002	MEX0003
Información al cierre de:	01/05/12	01/05/12	
Nombre	Empleado 1	Empleado 2	Empleado 3
ID People	MEX0000	MEX000	MEX00
Empresa	8 - MULTIASISTENCIA SERVICIOS, S.A. DE C.V.	21 - BBVA BANCOMER OPERADORA, S.A DE C.V.	
Reg nómina	87000	492	
Genero	F	M	
Edad	30.77	42.08	
Estado civil	Pareja de hecho	Soltero	
Número de hijos			
D.O.	MULTI	UCC	
Departamento	BC1201 - DETECCION DE FRAUDES OUT	BC1201 - DETECCION DE FRAUDES OUT	
Puesto	TW132 - EJECUTIVO ATN TELEFONICA B	WS79 - ASESOR TELEFO DETECCION FRAUDE	
Colectivo	TAA - Operativo (Confianza) Apoyo	OSS - Operativo (Sindicalizado)	
Antigüedad reconocida	5.21	18.10	
Antigüedad banco	5.21	18.10	
Antigüedad depto	3.49	0.36	
Antigüedad puesto	5.21	0.36	
Fecha de ingreso al grupo	12/03/07	21/04/94	
Fecha de rec antigüedad	12/03/07	21/04/94	
Año de valoración	2011	2011	
Catalogación	NORMAL		
Cumplimiento de objetivos		120.610	
Año de valoración	2010	2010	
Catalogación	NORMAL		
Cumplimiento de objetivos		156.040	
Año de valoración	2009	2009	
Catalogación	NORMAL		

2.4 Proyecto 3 - Hermes Global.

En este capítulo explicaré mi participación en el proyecto Hermes Global, el rol que desempeñe en este proyecto fue como líder de proyecto. Consistió en la instalación de una herramienta de Scoring que utilizaría la banca para aprobar o denegar créditos, esta aplicación automatizada ayudaría a los empleados de banca a tomar decisiones sobre si aprobar o denegar un préstamo, reduciendo la tasa de morosidad.

La presentación de esta herramienta se realizó en España y declarado como estratégico para la empresa. Derivado de las reuniones de usuarios. BP's y TI corporativos y homólogos locales, el proyecto se dividió en 3 fases: Aprovisionamiento, Motor de Calculo y Operativización. Los directores determinaron que a pesar que las fases eran dependientes podrían empalmar actividades con el fin de acortar la duración del proyecto global a un año.

Debido a que los planes respectivos para cada fase se fueron alargando, repleneando constantemente y sin un entregable terminado, el corporativo diseño una versión corta con un entregable parcial, pero de rápida generación, se me asigno como líder de este proyecto provisional, concluyéndolo en tiempo, forma y con el presupuesto acordado. Tiempo después el director de mi área me designa como nuevo líder del proyecto Global, por lo que tuve que realizar un cambio de actividades y equipo de trabajo de manera inmediata.

El verdadero reto de este proyecto fue reordenar actividades, desaparecer malas prácticas del personal que se me designo, y lo más importante eliminar la idea de proyecto interminable en el área de sistemas Holding.

Para este proyecto mi equipo de trabajo estaba formado por 6 consultores Oracle, 2 consultores Mainframe, 1 consultor Java, la duración del proyecto global fue aproximadamente de 3 años, en mi participación en este proyecto reduje el equipo a 3 consultores Oracle, 1 consultor Mainframe y 1 consultor Java, mi duración es este proyecto fue de 1 año, realizando la liberación en producción de todas las fases

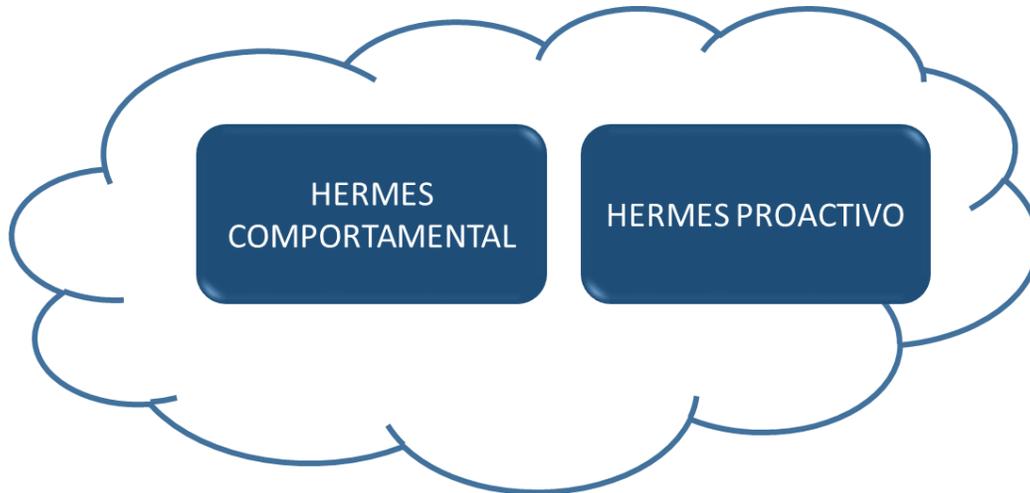
2.4.1 Antecedentes.

Uno de los principales objetivos del Área de Riesgos y de las Bancas minoristas, es el de disponer de un proceso de Scoring Proactivo que permita ofrecer límites de riesgo por cliente y producto con la granularidad requerida por las diferentes estrategias de negocio.

Dicha herramienta de software fue desarrollada por España e incluye los Modelos Proactivo y Comportamental, que componen Hermes Global, el Banco al ser una filial debía implementar esta herramienta corporativa.

En la figura 2.12 se muestra el concepto general de Hermes Global.

Figura 2.12 HERMES GLOBAL



El área usuaria de análisis financiero de riesgos fue la encargada de revisar con el corporativo la funcionalidad y alcance de esta herramienta, para lo cual se llevaron a cabo diversas conferencias y reuniones a lo largo de 8 meses, para el mejor entendimiento de la funcionalidad, así como las adaptaciones que debería aplicar para el país en particular, dado el comportamiento de su mercado, así como el volumen de clientes que maneja la institución. Estas reuniones se realizaron en conjunto con el área de gestión de iniciativas (BP's).

2.4.2 Objetivo

Implementar en el banco la herramienta comportamental para evaluar periódicamente la cartera de clientes mediante la aplicación del Scoring Proactivo a nivel cliente basado en la información histórica disponible del cliente, así como disponer de un proceso que proporcione límites de riesgo por cliente y producto, cubriendo las necesidades de las distintas estrategias de negocio que favorecerá la acción comercial y el adecuado control de riesgos.

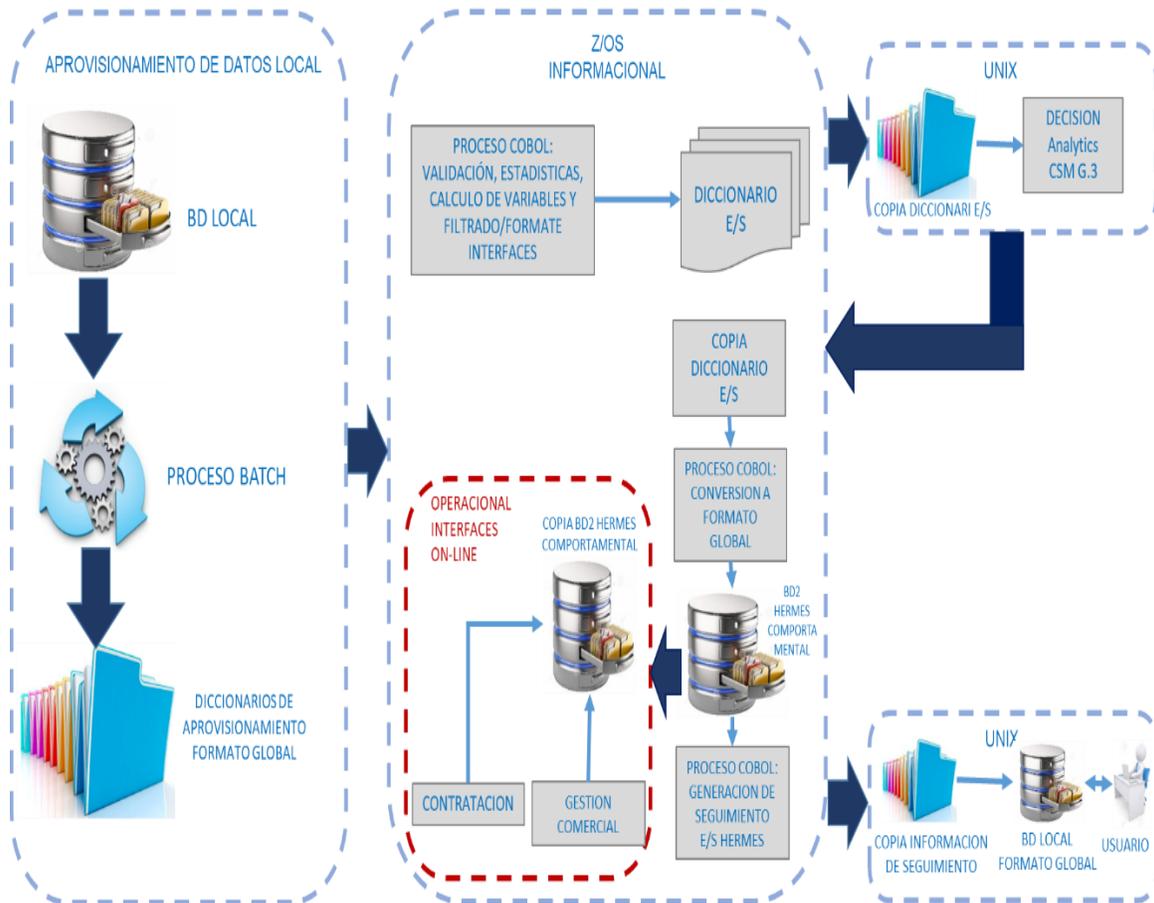
2.4.3 Descripción

Derivado de las reuniones que sostuvo el área del negocio con su homólogo del corporativo junto con el área de gestión de iniciativas y la dirección de sistemas Holding (subdirector), definieron el desarrollo en varias fases derivado de la naturaleza del proceso, las cuales se describen a continuación:

- Proceso multiplataforma.
- Volumetría de la información.
- Adaptaciones locales del proceso corporativo y viceversa.
- Infraestructura para su implantación.

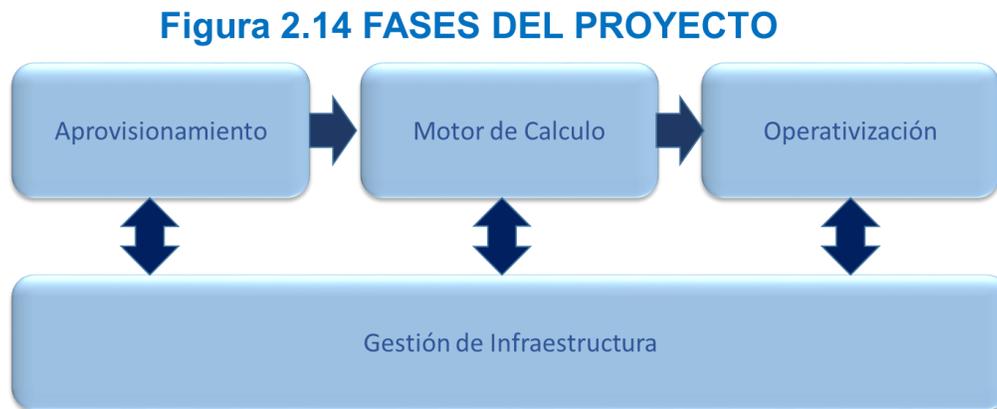
En la figura 2.13 se muestra flujo funcional de Hermes global:

Figura 2.13 FUNCIONAMIENTO HERMES GLOBAL



Al tratarse de una herramienta corporativa el proyecto se revisaba con el corporativo (áreas usuarias, BP's y TI corporativos y locales), las dudas, decisiones y pasos a seguir se tenían que corroborar y obtener su visto bueno para continuar, lo anterior debido a que no era permitido ningún cambio o suposición dentro de algún proceso de la herramienta, sin la revisión y aprobación previa del corporativo e incluía a todas las áreas participantes, este método de trabajo hacia más lento el desarrollo derivado a que dependíamos de los tiempos de respuesta del aplicativo, que también estaba revisando la implantación de esta herramienta en sus filiales.

En la figura 2.14 se muestran las grandes fases en que se dividió el proyecto.



2.4.4 Solución

Mi incorporación a este proyecto se realizó ya avanzado el desarrollo y con 3 planes ya definidos, por lo que mi primera tarea fue realizar reuniones técnicas con mi nuevo equipo para entender el punto donde nos ubicábamos, los pendientes, riesgos y atrasos de las actividades, así como los entregables de parte de las áreas de infraestructura (espacios, usuarios aplicativos, etc.). El atraso era considerable ya que se tenía un plan no actualizado desde 3 meses atrás, atraso en las entregas de espacio, usuarios y accesos por parte de las áreas de CCR, así como usuarios de negocio que tenían la idea de tomar a Hermes como fuente de información.

Una vez que entendí la situación específica del proyecto, decidí realizar juntas con mi equipo para la replanificación del plan de trabajo, el proyecto contaba con componentes instalados en producción. Sin embargo, estos no estaban liberados al área de gestión del servicio al ser una aplicación instalada de manera parcial y que no contaba con la aprobación por parte del usuario.

Dentro de las restricciones que nos marcaba la oficina de proyectos (área que daba seguimiento a los proyectos de TI), un plan se podía replanificar en tiempo, pero no en actividades, en caso de ser necesario se debería cancelar el proyecto y dar de alta uno nuevo con las actividades completas, adicional que cancelar un proyecto traía consigo una penalización en la evaluación anual de los participantes del mismo. Sin embargo, negocié con la oficina de proyectos se me permitiera realizar los cambios permitentes para poder finalizar el proyecto de la mejor manera. Junto con mi equipo de trabajo generamos los nuevos planes, los cuales tuvieron que ser planificados con actividades paralelas y no secuenciales como era la costumbre.

La ejecución del proceso se debería realizar de manera mensual, sin embargo, en su primera ejecución se tardó un mes y veinte días, sin contar con todos los componentes instalados, por lo que adicional a realizar la instalación de los componentes faltantes, tendríamos que realizar una reingeniería a los componentes instalados con el objetivo de reducir los tiempos de ejecución, lo cual implicaba un doble trabajo por parte de todo el equipo.

A continuación, explicaré el desarrollo de los bloques en que se dividió el proyecto.

Fase 1: Aprovisionamiento

El aprovisionamiento debía contener todas las cuentas de pasivo y activo del universo de clientes que cumplan con las condiciones especificadas, conformado por personas físicas (naturales), mayores de edad, no empleados en la Entidad Financiera y su consolidación debería ser a nivel de cliente.

- Pasivo Vigente: pasivos vigentes a la fecha de referencia, incluidas cuentas corrientes bloqueadas, embargadas, etc., plazos fijos vencidos con saldos pendientes de pago y cuentas corrientes (con o sin acuerdos, con saldo deudor y/o acreedor).
- Activo Vigente: activos vigentes a la fecha de referencia y con vencimiento posterior a dicha fecha, o bien, con vencimiento anterior a la fecha de referencia, pero con pagos pendientes a la fecha de referencia.
- Activo No Vigente: activos con fecha de vencimiento en los dos años anteriores a la fecha de referencia, o bien, con vencimiento anterior a dos años antes de la fecha de referencia, pero con pagos pendientes a esta fecha. En ambos casos, sin pagos pendientes a la fecha de referencia. Los incobrables se consideran no vigentes.

En la figura 2.15 se muestra el mapa del tipo de variables que contuvo el aprovisionamiento:

Figura 2.15 VARIABLES DEL APROVISIONAMIENTO

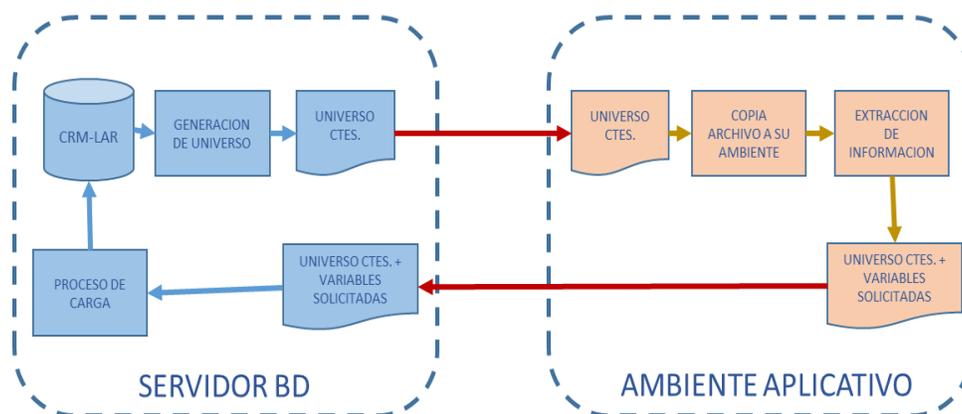


Dentro del análisis inicial (antes de mi asignación al proyecto) el equipo anterior encontró que la información para el aprovisionamiento se encontraba almacenada en su mayoría en el Customer Relationship Management (CRM) del Laboratorio Analítico Regional (LAR), para algunas variables que no estaban contenidos en dicha BD se solicitó su aprovisionarlas de los aplicativos fuente, estas solicitudes fueron atendidas por cada área a las que se les solicito, sin embargo nunca se incluyeron en el aprovisionamiento (solo eran interfaces que llegaban al servidor pero no se cargaban a la base de datos, algunas interfaces incluso llegaban vacías), por lo que una de las actividades de la replanificación en mi equipo de trabajo fue analizar estas interfaces e incluirlas en aprovisionamiento o solicitar a los aplicativos la revisión de sus interfaces, uno de los problemas a los que me enfrente en esta situación fue que los aplicativos contaban con el VoBo de sus interfaces y por lo tanto no contábamos con presupuesto adicional para solicitar un cambio y tampoco pudimos presentarlo como una incidencia en el proceso, dado que ya habían pasado meses desde su instalación, así también dispuse las actividades para analizar los programas de las extracciones ya diseñadas para encontrar errores y/o reducir el tiempo de procesamiento, en esta fase replanificada nos encontramos con muchos errores, los formatos de aprobación que no debieron haberse otorgado, sin embargo este análisis nos sirvió bastante para poder continuar con el proyecto y reducir tiempos de procesamiento.

Después del análisis continuamos con el diseño de los componentes, algunas variables eran directas y para otras eran necesario realizar cálculos para su generación. El flujo del aprovisionamiento consistía en generar un universo de clientes basado en las reglas del negocio, este universo era enviado a cada aplicativo, ellos lo subían a su ambiente productivo y realizaba la extracción de los campos solicitados, una vez que se tenía esta interfaz completa, era enviada de regreso al servidor para realizar el proceso de carga a la base de datos del Laboratorio Analítico Regional (LAR).

En la figura 2.16 se muestra el diseño del flujo de estas interfaces solicitadas.

Figura 2.16 FLUJO INTERFACES DE APLICATIVOS



Dado que tardaban días los aplicativos en regresar las interfaces solicitadas, pudimos encontrar que no era necesario esperar a tener todas las variables para iniciar algunos cálculos, por lo que en la fase de diseño mi equipo y yo encontramos una oportunidad de adelantar estos procesos, esto nos ayudó a reducir el tiempo de procesamiento, modificando la secuencia de los procesos diseñados inicialmente.

En la fase de construcción mi equipo se enfocó en generar los componentes faltantes y realizar la reingeniería de los procesos existentes, una vez concluida esta tarea se llevaron a cabo las pruebas funcionales y al terminar las pruebas de usuario, para poder conseguir que el usuario volviera a validar, realice negociaciones con ellos indicándoles que era necesario dado que se modificaron los componentes para la reducción del tiempo de procesamiento. Las actividades realizadas por mi parte se enfocaron al seguimiento puntal de las actividades, convocar a reuniones con aplicativos y/o usuarios, negociar con las áreas de infraestructura accesos y espacios, y diferentes actividades pendientes.

En la implementación mi equipo de trabajo y yo logramos el aprovisionamiento de todas las 742 variables, se generaron los 14 diccionarios solicitados para el motor, cabe señalar que durante el transcurso de este desarrollo hubo pequeños proyectos derivados de los cambios en las reglas de negocio que se fueron dando con el tiempo y los cambios en el mercado.

En la tabla 1.1 se muestran las interfaces generadas que se denominaron diccionarios en este proyecto.

Tabla 1.1 LISTA DICCIONARIOS HERMES

DICCIONARIO	DESCRIPCION	TOTAL DE CAMPOS
CRIHC001	DICCIONARI DE CLIENTES	45
CRIHC002	DICCIONARIO PROACTIVO CLIENTES	72
CRIHC003	DICCIONARIO COMPORTAMENTAL DE CLIENTES	30
CRIHC004	DICCIONARIO 1+2+HISTORICO DE CONTRATOS DE ARJETAS	69
CRIHC005	DICCIONARIO CUENTAS	102
CRIHC006	DICCIONARIO PRESTAMOS	67
CRIHC007	DICCIONARIO PRESTAMOS HIPOTGECARIOS	58
CRIHC008	DICCIONARIO OTRAS CUENTAS DE PASIVO	28
CRIHC009	DICCIONARIO VALORES	19
CRIHC010	DICCIONARIO SEGUROS	21
CRIHC011	DICCIONARIO MORAE E INVERSION IRREGULAR	37
CRIHC012	DICCIONARIO VINCULACION	89
CRIHC013	DICCIONARIO LINEAS DE CREDITO	68
CRIHC014	DICCIONARIO DE DATOS DE CIRBE	37
		742

La duración del proyecto fue aproximadamente de 8 meses, en la tabla 1.2 se describen las características del entorno central utilizado en el desarrollo.

Tabla 1.2 Host

Entorno HOST	
Sistema Operativo	MVS
Lenguaje de Programación	COBOL
Base de Datos	DB2
Transacciones On-Line	CICS
Monitor Transaccional	IMS

En la tabla 1.3 se describen las características del entorno distribuido utilizado.

Tabla 1.3 Distribuido

Entorno Distribuido	
Sistema Operativo	UNIX
Lenguaje de Programación	Lenguaje C, JAVA, SASbase

Fase 2 - Motor de cálculo

El motor de cálculo es una herramienta proactiva que permite evaluar mensualmente el comportamiento de personas físicas y físicas con actividad empresarial mediante la aplicación del Scoring Proactivo a nivel cliente basado en la siguiente información:

- Datos transaccionales.
- Flujo de efectivo.
- Tenencia de productos.

Obteniendo así la capacidad de pago mensual lo que permite generar límites máximos para los siguientes productos:

- Auto (finanza y bancaria).
- Consumo (Nómina, PPI y Anticipo de Nómina).
- Tarjeta de Crédito (finanza y bancaria).
- Hipotecario.

Los datos generados por el motor eran únicos y re-aprovechables para diferentes áreas de la empresa.

Los componentes del aplicativo se dejaron instalados en el servidor Host de Producción México, por lo que fue necesario habilitar acceso a la máquina local para el equipo de Hermes Global España y crear usuarios de TSO (Time Sharing Option) con los permisos necesarios para realizar esta instalación.

Se creó una librería por cada tipo de componente, en la tabla 1.4 se muestra en detalle.

Tabla 1.4 LISTA DE LIBRERIAS

Componente	Librería de instalación
COPYS	lbpais.COPY
MAPAS	lbpais.MAPAS
Programas Batch y rutinas	lbpais.SRC
Transacciones	lbpais.SRL
JCL's	lbpais.JCL
DCLGEN's	lbpais.DCLGEN
Componentes de librería EX.EXPDS111.PUNLOAD.BMC	lbpais.PUNLOAD

lbpais: se corresponde con la nomenclatura propia de la instalación de cada país.

En la ejecución del primer proyecto del motor se instalaron 346 componentes en el ambiente Host enviados por el corporativo, en la tabla 1.5 se enlista el detalle.

Tabla 1.5 INVENTARIO DE COMPONENTES

Componente	Total
COPYS	
De comunicación del aprovisionamiento local	14
De programas batch y rutinas	60
De transacciones	5
De comunicación entre módulos	4
TABLAS	20
DCLGEN's	20
COMPONENTES DE LA LIBRERÍA	13
MAPAS	5
PROGRAMAS	
BATCH	85
ON-LINE (transacciones)	5
Módulos (rutinas)	14
PROCESOS BATCH's	101

Muchos de los componentes instalados tuvieron que ser tipificados para México, ya que los originales contenían especificaciones de la arquitectura de España, así también muchas de las variables del aprovisionamiento se generaron con una inicialización especial por ser productos no existentes en México.

También se instaló la herramienta de Experian conocido como DA (Decision Analytics), en la cual los campos numéricos no informados, se deberían diferenciar de los que tengan valor cero. Esta herramienta sirve para manejar límites propuestos por producto, para cada cliente. Por lo anterior, SDS (herramienta del DA), debería estar conectado a los aplicativos del Banco, en este caso Terminal Financiero, para que los ejecutivos cuenten con opciones de oferta hacia los clientes.

En cada país se generaron localmente los archivos de Estrategias, uno para Proactivo Clientes (PPBBVAPC.ser) y otro para Comportamental Tarjetas (PPBBVATJ.ser).

Una vez finalizados los procesos globales que generan los archivos con datos de Proactivo Cliente y Comportamental Tarjetas para la ejecución de la herramienta, fue necesario transmitir ambos archivos. Todos estos componentes se instalaron localmente en México, debiéndose adaptar a la nomenclatura específica del país y modificar lo necesario, según las características de la arquitectura local. Se gestionó la implantación en cada entorno: Desarrollo, Integración y Producción, se realizaron las pruebas que correspondan en cada caso. Todo lo anterior se generó durante el primer plan del proyecto, debido a que esta parte fue supervisada puntualmente con España la instalación de los componentes fue correcta, sin embargo, existían muchos procesos que tenían la finalización forzada para su término correcto (Force OK), lo anterior se realizó debido a que todas las aplicaciones eran monitoreadas por el área de Producción y cada terminación errónea era contabilizada afectando los niveles de cumplimiento y efectividad de cada dirección.

El reto al tomar este proyecto y replanificarlo fue revisar con mi equipo toda la cadena de procesos instalados y por instalar, el equipo de España mantenía reuniones semanales con el equipo de desarrollo en México, tratando de que la diferencia de horario entre ambos países no afectará se acordaron las reuniones con un horario de 9 a 10:30 hora de México, una de las cuestiones más importantes que tuve que aprender fueron las políticas del ambiente central (Mainframe) dado que nunca había trabajado en este ambiente y por lo tanto no tenía conocimientos previos que me ayudaran en este proceso. Sin embargo, conté con el apoyo de mi equipo de trabajo, escuché sus recomendaciones, confié en su experiencia y logramos replanificar el proyecto de la mejor manera, existía el 20% de componentes con fallo o con ejecución congelada de manera temporal, este desarrollo lo realizamos con actividades en paralelo de análisis, diseño, desarrollo y pruebas funcionales, la división de componentes se realizó por bloques de la cadena de ejecución, así como las implantaciones.

Mis actividades en este proyecto fueron de supervisión de las actividades replanificadas, gestiones de espacio y seguridad de los datos con las áreas de infraestructura, reuniones con los usuarios locales y el área de Desarrollo España, así como negociaciones de pruebas de usuario con las áreas del negocio. Uno de

los mayores riesgos a los que nos enfrentamos como equipo de Desarrollo fueron los tiempos de respuestas por parte de Desarrollo España, así como de las áreas locales, por ser temas revisados anteriormente (primer plan del proyecto).

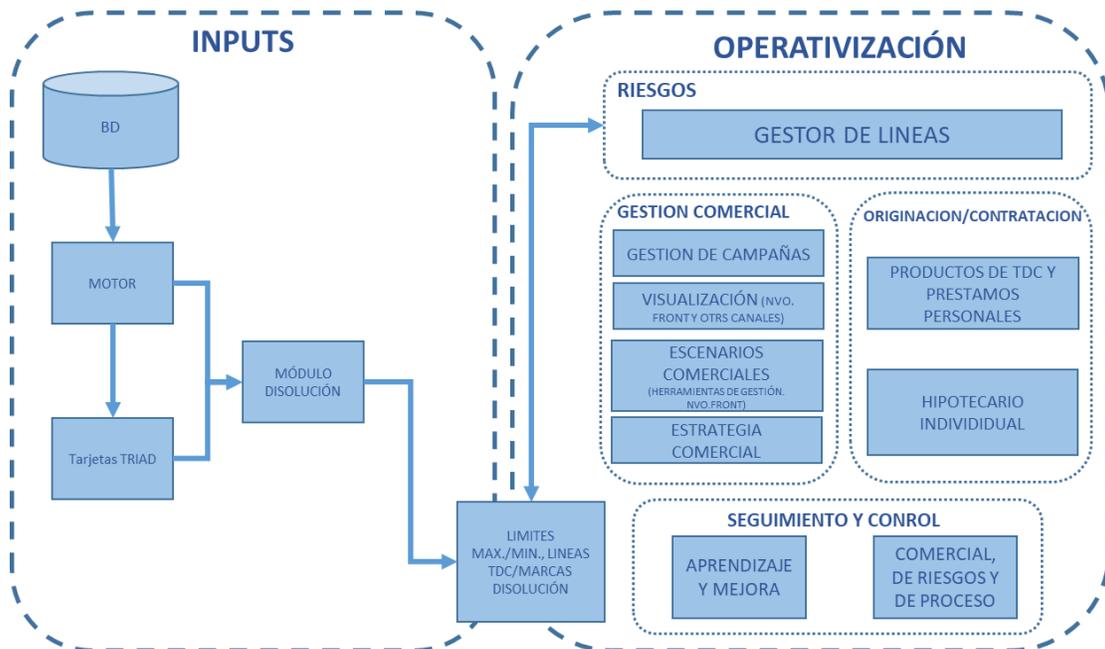
La salida del motor consistió en 2 interfaces uno con 2750 variable de salida y otra para la siguiente fase La Operativización con los límites de créditos calculados, la duración de los proyectos implicados en esta fase tuvo una duración aproximada de 10 meses.

Fase 3 - Operativización

La Operativización de Hermes Proactivo contemplo a personas físicas y físicas con actividad empresarial, lo cual permitirá ofrecer al banco una herramienta capaz de proveer información actualizada de los límites de crédito de los clientes para los productos de Auto, Consumo, Tarjeta de Crédito e Hipotecario, mantenimiento de líneas de crédito de TDC, así como también soluciones de recuperación, permitiendo así, agilizar y estructurar los procedimientos comerciales y de gestión, con la finalidad de mejorar los circuitos operativos de asesoramiento y análisis de los clientes y sus operaciones.

En la figura 2.17 se muestra la conceptualización de la Operativización.

Figura 2.17 OPERATIVIZACIÓN



Se creó la interfaz para la comunicación Batch entre el motor Hermes y el Gestor de Líneas, para tener actualizados los Límites Máximos proveídos por la herramienta motor Hermes.

La interfaz se cargaba mensualmente para que Hermes alimente al gestor de líneas con la evaluación mensual mediante la aplicación del Scoring, esta interfaz contiene los límites máximos de los siguientes productos:

- Auto
- Consumo (nómina y PPI)
- Tarjeta de crédito (finanzia y banco)
- Hipotecario

El proceso de Operativizacion realiza los cálculos necesarios para la administración de los límites disponibles, tomado en cuenta factores como la morosidad, cancelación de relación de clientes, fallecidos, exclusión de clientes, desestimación de campañas, los cuales son determinantes para calcular lo dispuesto por el cliente en los productos de activo que se encuentren en un proceso de contratación, pero que aún no han sido formalizados o campañas ofertadas (aun y cuando algunos de estos productos no estén definidos en la salida del motor Hermes Proactivo). Con lo cual se estima la capacidad de endeudamiento máximo que el cliente puede adquirir con la Entidad para repartir entre los distintos productos.

Este límite máximo será asignado por el Motor Hermes Proactivo de manera mensual considerando datos transaccionales, tenencia de productos, capacidad de pago, etc. Los límites estarán dados en la siguiente agrupación:

- A nivel General de Cliente.
- A nivel de Tipo de Riesgo.
 - Tipo de riesgo Real.
 - Tipo de riesgo con Garantía.
 - Tipo de riesgo Personal.
- A nivel de Agrupación de Riesgos – Producto.
 - Auto.
 - Consumo.
 - Tarjeta de Crédito.
 - Hipotecario.

Al tomar el proyecto bajo mi cargo, me encontré que este bloque del proyecto se encontraba sin seguimiento, con un plan no actualizado hacía más de 6 meses, con actividades totalmente abandonadas por el anterior líder y con un avance bastante pobre y no real plasmado en el plan, una vez que entendí el punto en donde nos encontrábamos, mi equipo y yo nos enfocamos a replanificar el plan, en este caso las fechas fueron cambiadas, las actividades necesarias estaban contempladas por lo que no se agregaron actividades adicionales, realice varias reuniones con los

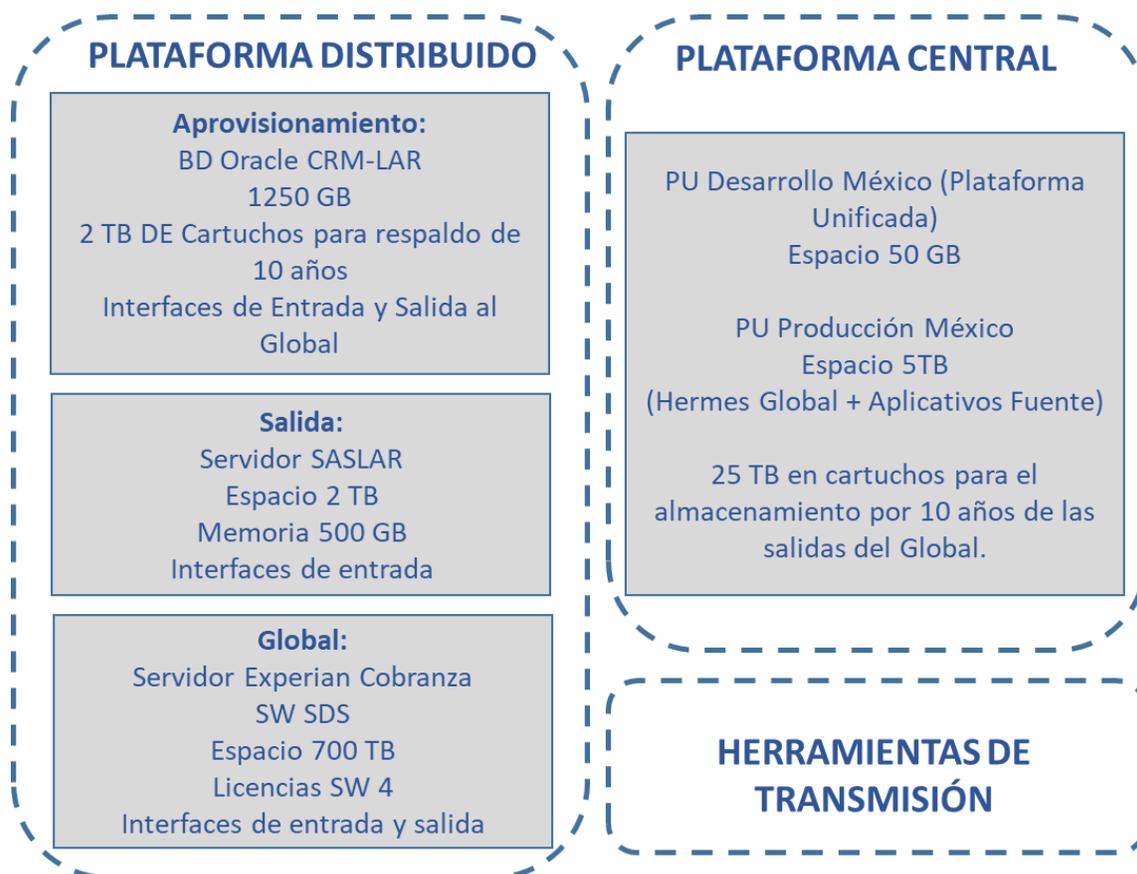
diferentes aplicativos, así como con las áreas de CCR para negociar los tiempos de las actividades pendientes y lograr la replanificación.

A pesar que el plan de desarrollo se encontraba plasmado de manera secuencial, tuvimos que realizar el análisis, diseño, construcción, pruebas funcionales y pruebas de usuario de manera paralela, incluso algunas implantaciones, dividiendo en bloques las diferentes áreas involucradas en la Operativización.

Fase - Gestión infraestructura

A lo largo del proyecto se tuvieron que realizar diferentes gestiones de infraestructura, en la figura 2.18 se muestra el mapa conceptual de lo solicitado durante todo el proyecto.

Figura 2.18 INFRAESTRUCTURA



En el Banco existe el área de CCR (Centro Corporativo Regional), la cual es la dirección encargada de la atención a los temas de infraestructura, comunicación, servidores, seguridad, etc.

Para la gestión de recursos se generó un documento de entendimiento donde se plasman los siguientes datos:

- Objetivos del Negocio.
- Alcance de la solicitud.
- Beneficios cuantitativos y cualitativos.
- Situación Actual
- Solución conceptual
- Flujo del aplicativo a instalar.
- Productos de SW a adquirir
- Infraestructura solicitada

Para el desarrollo del proyecto se solicitaron 5 peticiones de infraestructura debido al ejercicio presupuestal que manejaba la empresa. Al tomar el proyecto a mi cargo como líder de proyecto, tuve que revisar todas y cada una de las entregas pactadas con las áreas de CCR, descubriendo que muchas de estas entregas tenían un retraso considerable, así como falta de atención por parte de estas áreas. La justificación de las mismas fue que el anterior líder no respondió nunca a los correos de alertamiento enviados por ellos, por lo anterior tuve que retomar cada uno de los temas pidiendo al anterior líder el reenvío de dichos correos, para generar reuniones con cada área específica, así como negociaciones para la entrega de los productos pendientes.

La mayoría de estos pendientes fueron espacio, accesos y ambientes productivos con niveles de servicio. Por ejemplo, la instalación en tiempo del DA (plan inicial), este se instaló en un área de desarrollo y no de producción, para realizar el cambio correspondiente, fue necesario solicitar presupuesto adicional y poder realizar el cambio.

2.4.5 Resultados

Una vez terminados todos los planes e implantada toda la cadena del proceso, Hermes se tomó como base en la decisión de admisión para cualquier tipo de cliente. Resultando una herramienta muy útil para la selección de campañas, incrementos de líneas, precios, así como una llave más de decisión para estrategias de autorizaciones y cobranza.

A continuación, se describen los resultados particulares de Hermes que se obtuvieron a nivel negocio.

A nivel cliente:

- Calificación de riesgo por cliente
- Asignación de límites máximos, plazos, condiciones, etc. por producto.
- Incorporación en el proceso de admisión sin diferenciar canales (sucursales, canales externos)

En los procesos de admisión:

- Menor cantidad de documentación en el caso de los clientes de la institución
- Nuevas Campañas

Gestión de políticas de riesgos

- Homogeneización de criterios.
- Nuevos modelos de riesgo para los clientes del banco.
- Flexibilidad en la parametría.

Pre-autorizaciones

- Al contar con una herramienta de parametría e información integrada, se pueden incluir ofertas permanentes para los clientes
- Multiproducto

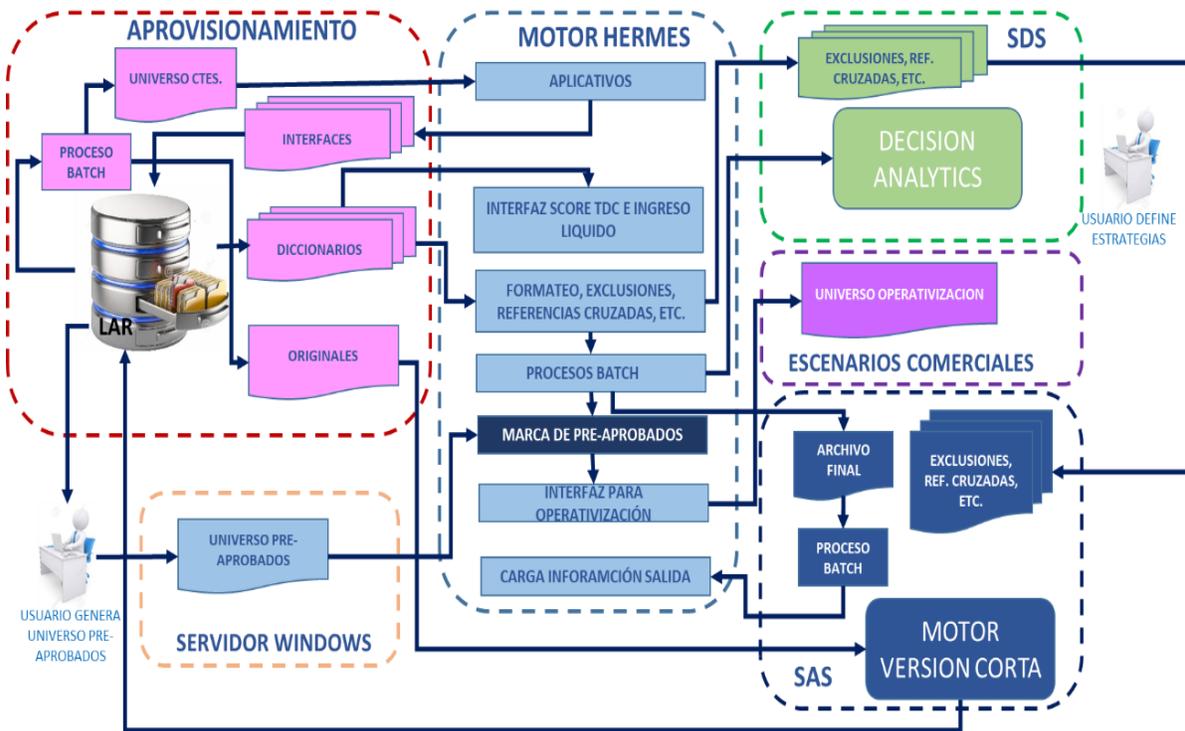
Información operativa

- Posibilidad de integrar información comercial como propensión de un cliente a tomar una oferta o a cancelar un producto

La cadena implementada mantuvo el intercambio de interfaces entre las plataformas distribuida y Central. El flujo del proceso iniciaba con el aprovisionamiento de la información del universo de clientes, este universo era enviado a los diferentes aplicativos (la mayoría en Mainframe), regresando las interfaces a la BD para la generación de los diccionarios que fungían como entrada al motor de cálculo, los diccionarios también eran entrada para las aplicaciones de Score TDC e Ingreso líquido, por su parte el motor iniciaba su ejecución con el proceso Formateo, Exclusiones, Referencias Cruzadas, las interfaces de salida generadas eran transferidas al servidor SDS, para ser transferidas al servidor SAS, cabe señalar que el servidor SDS solo era un puente de transmisión hacia el servidor SAS, debido a que no se consideró en el presupuesto la instalación de canales de comunicación desde el Mainframe al servidor SAS, una vez que los archivos eran depositados en SAS, el usuario llevaba a cabo revisiones de los mismos. Por su parte el motor

continuaba con su cadena de ejecución, al finalizar este bloque, por una parte, se ejecutaba la marca de clientes pre-aprobados, esta marca se realizaba mediante el envío de un archivo generado por el usuario de riesgos a un servidor Windows, el cual era tomado por medio de un proceso automático para transmitirlo al Mainframe y realizar el proceso de marcado. Por otra parte, las salidas eran enviadas al servidor SDS para la ejecución del Decision Analytics, en este punto el usuario tenía 3 días para ejecutar las reglas de negocio por medio de esta herramienta, una vez concluida esta tarea la interfaz de salida era enviada al siguiente paso de la cadena. Como paso final se realizaba la generación de la interfaz de Operativización, esta interfaz era transmitida al ambiente de Escenarios Comerciales, en la cual se contenían los límites de crédito, líneas de TDC y/o marcas de disolución que eran consultadas mediante transacciones línea por los gestores de las oficinas. En la figura 2.19 se muestra de manera gráfica la solución implementada.

Figura 2.19 SOLUCIÓN PROYECTO HERMES



El tiempo de ejecución de Hermes paso de un mes y veinte días a veintiocho días de inicio a fin, sin embargo debido a las fallas presentadas a lo largo de los 3 años de duración del proyecto, existía desconfianza por parte de las áreas de negocio y con una versión corta funcionando casi 2 años, fue necesario que a nivel dirección se tomara la decisión del apagado de la versión corta para obligar a todas las áreas de negocio utilizar una herramienta corporativa que costó mucho tiempo, dinero y esfuerzo concluir.

Por otra parte, durante casi un año fungimos mi equipo y yo como soporte de la aplicación, dado que el área de Gestión del Servicio intento en 3 ocasiones tomar el proceso sin éxito, por lo que se realizó la delegación del mismo por partes.

En este proyecto me enfrente por primera vez a reuniones a nivel dirección donde mi labor fue básicamente justificar los tiempos y el trabajo pendiente, también me dio la satisfacción de sacar adelante un proyecto considerado interminable dentro de la empresa.

Capítulo 3 – Marco Teórico Minería de Datos.

A continuación, se explicará el marco teórico de la minería de datos, derivado que el proyecto principal expuesto en este reporte que es el diseño e implementación de un Modelo de Minería.

Conocido algunas veces como "descubrimiento de conocimientos en bases de datos", el término "minería de datos" (SAS, 2019) no se acuñó sino hasta la década de 1990. Pero su base comprende tres disciplinas científicas entrelazadas:

- Estadística - El estudio numérico de relaciones de datos.
- Inteligencia Artificial - Inteligencia similar a la humana exhibida por software y/o máquinas.
- Aprendizaje Automático - Algoritmos que pueden aprender de datos para hacer predicciones.

Lo que era antiguo es nuevo otra vez, ya que la minería de datos continúa evolucionando para igualar el ritmo del potencial sin límites del big data y poder de cómputo asequible.

En la figura 3.1 se muestra gráficamente las disciplinas entrelazadas en la minería de datos.

Figura 3.1 DISCIPLINAS EN LA MINERIA DE DATOS



En la última década, los avances en el poder y la velocidad de procesamiento nos han permitido llegar más allá de las prácticas manuales, tediosas y que toman mucho tiempo al análisis de datos rápido, fácil y automatizado. Cuanto más complejos son los conjuntos de datos recopilados, mayor es el potencial que hay para descubrir insights relevantes. Los comerciantes detallistas, bancos, fabricantes, proveedores de telecomunicaciones y aseguradoras, entre otros, utilizan la minería de datos para descubrir relaciones entre todas las cosas, desde precios, promociones y demografía hasta la forma en que la economía, el riesgo, la competencia y los medios sociales afectan sus modelos de negocios, ingresos, operaciones y relaciones con clientes.

3.1 ¿Qué es la Minería de Datos?

La minería de datos (Duncan, 2018) es el proceso de detección de la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

Estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos. Los modelos de minería de datos se pueden aplicar en escenarios como los siguientes:

- Pronóstico: cálculo de las ventas y predicción de las cargas del servidor o del tiempo de inactividad del servidor.
- Riesgo y probabilidad: elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- Recomendaciones: determinación de los productos que se pueden vender juntos y generación de recomendaciones.
- Búsqueda de secuencias: análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
- Agrupación: distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades.

3.2 ¿Cómo funciona la Minería de Datos?

Derivado de que el volumen de datos producidos se duplica cada dos años. Los datos no estructurados por sí solos conforman el 90% del universo digital. Pero más información no significa necesariamente más conocimientos

La minería de datos le permite:

- Filtrar todo el ruido caótico y repetitivo en sus datos.
- Entender qué es relevante y luego hacer un buen uso de esa información para evaluar resultados probables.
- Acelerar el ritmo de la toma de decisiones informadas.

La minería de datos es una piedra angular de la analítica, la cual le ayuda a desarrollar los modelos que pueden descubrir conexiones dentro de millones o miles de millones de registros.

La minería de datos, como una disciplina compuesta, representa diversos métodos o técnicas que se utilizan en diferentes capacidades analíticas que abordan una gama de necesidades organizacionales, hacen diferentes tipos de preguntas y utilizan diferentes niveles de aportación humana o reglas para llegar a una decisión.

En la base de la minería de datos existen 3 tipos de modelado:

- Descriptivo
- Predictivo
- Prescriptivo.

A continuación, se detalla cada tipo de modelado.

3.2.1 Modelado Descriptivo

Descubre similitudes o agrupaciones compartidas en datos históricos para determinar razones detrás del éxito o el fracaso, como la clasificación de clientes por preferencias de productos o sentimiento. Algunas técnicas de ejemplo incluyen:

- Agrupación de clústeres: Agrupación de registros similares.
- Detección de anomalías: Identificación de valores atípicos multidimensionales.
- Aprendizaje de reglas de asociación: Detección de relaciones entre registros.
- Análisis de componentes principales: Detección de relaciones entre variables.
- Agrupación por afinidad: Agrupación de personas con intereses comunes o metas similares.

3.2.2 Modelado Predictivo

Este modelado llega más a fondo para clasificar eventos en el futuro o calcular resultados desconocidos – por ejemplo, el uso de evaluación de crédito para determinar la probabilidad de que una persona pague un préstamo. El modelado predictivo también ayuda a descubrir insights de cosas como la rotación de clientes, respuesta a campañas o coberturas por impago de créditos. Algunas técnicas de ejemplo incluyen:

- Regresión: Medida de la solidez de la relación entre una variable dependiente y una serie de variables independientes.
- Redes Neuronales: Programa de computadora que detecta patrones, hacen predicciones y aprenden.
- Árboles de Decisión: Diagrama en forma de árbol en los que cada rama representa una incidencia probable.
- Máquinas Vectoriales de Soporte: Modelos de aprendizaje supervisado con algoritmos de aprendizaje asociado.

3.2.3 Modelado Prescriptivo

Con el incremento de los datos no estructurados de la Web, campos de comentarios, libros, correo electrónico, PDFs, audio y otras fuentes de texto, la adopción de la minería de texto como disciplina relacionada con la minería de datos también ha crecido de manera considerable. Necesita la posibilidad de analizar, filtrar y transformar con éxito datos no estructurados para incluirlos en modelos predictivos para mejorar la precisión de las predicciones.

El modelado prescriptivo se centra en variables y restricciones internas y externas para recomendar uno o más cursos de acción. Algunas técnicas de ejemplo incluyen:

- Analítica predictiva más reglas: Creación de reglas que pasaría si/entonces a partir de patrones y predicción de resultados.
- Optimización del marketing: Simulación de la mezcla de medios más ventajosa en tiempo real para lograr el más alto ROI posible.

Al final, no se debe ver la minería de datos como una entidad independiente porque el pre procesamiento (preparación y explotación de datos) y el pos procesamiento (validación de modelos, calificación y monitoreo del desempeño de modelos) son igualmente esenciales.

3.3 Ciclo de la Minería de Datos

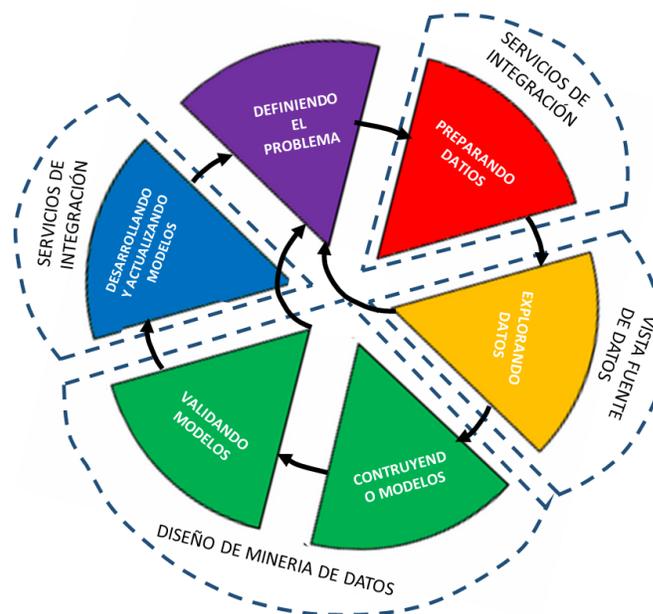
La generación de un modelo de minería de datos forma parte de un proceso mayor (Duncan, 2018) que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo. Este proceso se puede definir mediante los siguientes seis pasos básicos:

1. Definir el problema
2. Preparar los datos
3. Explorar los datos
4. Generar modelos
5. Explorar y validar los modelos
6. Implementar y actualizar los modelos

El proceso es cíclico, lo que significa que la creación de un modelo de minería de datos es un proceso dinámico e iterativo. Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de minería de datos adecuados y que, por tanto, debe buscar más datos. O bien, puede generar varios modelos y descubrir entonces que no responden adecuadamente al problema planteado cuando los definió y que, por tanto, debe volver a definir el problema. Es posible que deba actualizar los modelos una vez implementados debido a que haya más datos disponibles. Puede que haya que repetir cada paso del proceso muchas veces para crear un modelo adecuado.

En la figura 3.2 se ilustra gráficamente los pasos, así como su relación.

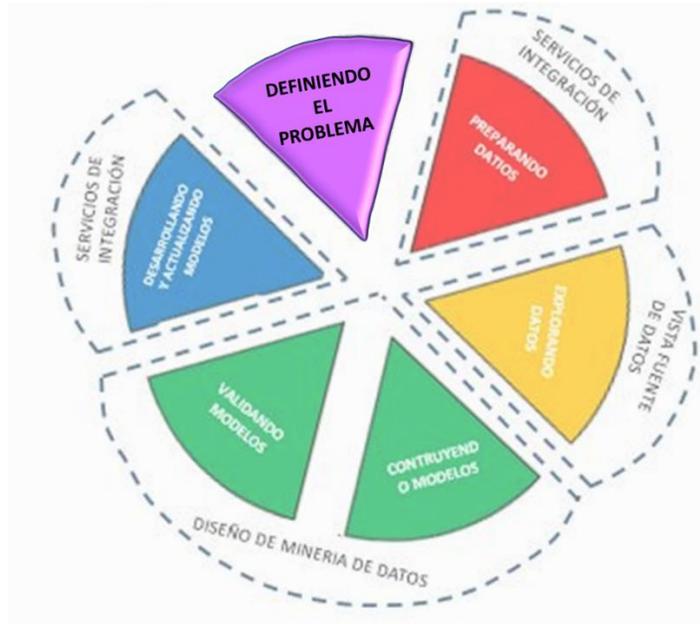
Figura 3.2 PASOS MODELO DE MINERIA



3.3.1 Definir Problema.

El primer paso del proceso de minería de datos, tal como se resalta en la figura 3.3, consiste en definir claramente el problema y considerar formas de usar los datos para proporcionar una respuesta para el mismo.

Figura 3.3 DEFINIR PROBLEMA



Este paso incluye analizar los requisitos empresariales, definir el ámbito del problema, definir las métricas por las que se evaluará el modelo y definir los objetivos concretos del proyecto de minería de datos. Estas tareas se traducen en preguntas como las siguientes:

- ¿Qué está buscando? ¿Qué tipos de relaciones intenta buscar?
- ¿Refleja el problema que está intentando resolver las directivas o procesos de la empresa?
- ¿Desea realizar predicciones a partir del modelo de minería de datos o solamente buscar asociaciones y patrones interesantes?
- ¿Qué resultado o atributo desea predecir?
- ¿Qué tipo de datos tiene y qué tipo de información hay en cada columna? En caso de que haya varias tablas, ¿cómo se relacionan? ¿Necesita limpiar, agregar o procesar los datos antes de poder usarlos?
- ¿Cómo se distribuyen los datos? ¿Los datos son estacionales? ¿Los datos representan con precisión los procesos de la empresa?

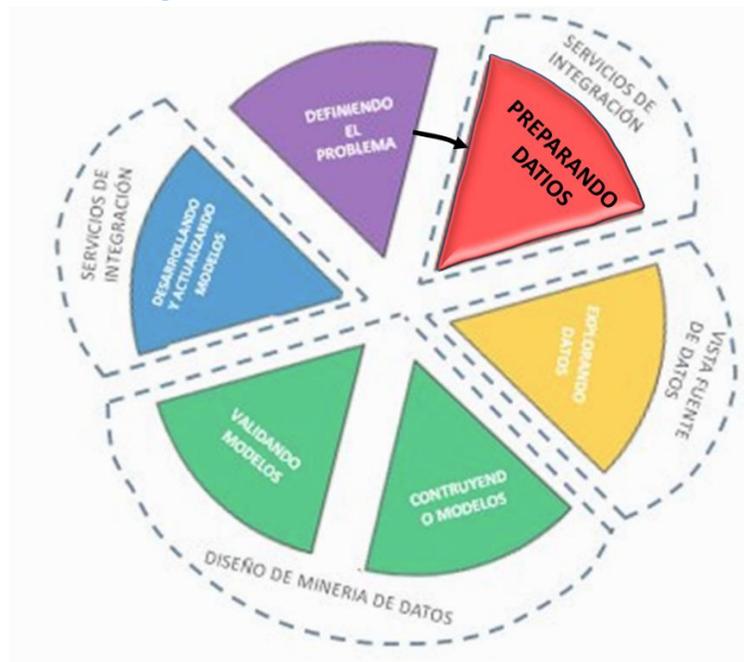
Para responder a estas preguntas, puede que deba dirigir un estudio de disponibilidad de datos para investigar las necesidades de los usuarios de la empresa con respecto a los datos disponibles. Si los datos no abarcan las necesidades de los usuarios, podría tener que volver a definir el proyecto.

También debe considerar las maneras en las que los resultados del modelo se pueden incorporar en los indicadores de rendimiento clave (KPI) que se utilizan para medir el progreso comercial.

3.3.2 Preparar Datos

El segundo paso del proceso de minería de datos, como se indica en la figura 3.4, consiste en consolidar y limpiar los datos identificados en el paso Definir el problema.

Figura 3.4 PREPARAR DATOS



Los datos pueden estar dispersos en la empresa y almacenados en formatos distintos; también pueden contener incoherencias como entradas que faltan o incorrectas. A continuación, se enlistan algunas consideraciones que se deben tener en cuenta en este paso:

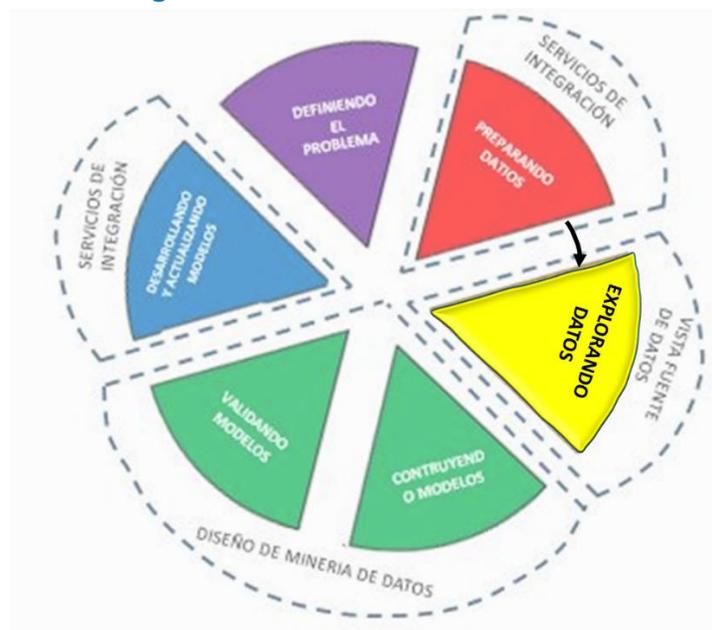
- Datos Malos o poco útil
 - Incompletos
 - Con ruido
 - Inconsistentes
 - Duplicados
- Datos de Calidad
 - Recuperar información incompleta
 - Eliminar outliers (Sanjuán, economipedia, s.f.)
 - Resolver conflictos/inconsistencias

La limpieza de datos no solamente implica quitar los datos no válidos o interpolar valores que faltan, sino también buscar las correlaciones ocultas en los datos, identificar los orígenes de datos que son más precisos y determinar qué columnas son las más adecuadas para el análisis. Por ejemplo, ¿debería utilizar la fecha de envío o la fecha de pedido? ¿Qué influye más en las ventas, la cantidad, el precio total o un precio con descuento? Los datos incompletos, los datos incorrectos y las entradas que parecen independientes, pero que de hecho están estrechamente correlacionadas, pueden influir en los resultados del modelo de maneras que no espera.

3.3.3 Explorar datos

El tercer paso del proceso de minería de datos, como se muestra en la figura 3.5, consiste en explorar los datos preparados.

Figura 3.5 EXPLORAR DATOS



Debe conocer los datos para tomar las decisiones adecuadas al crear los modelos de minería de datos. Entre las técnicas de exploración se incluyen:

- Calcular los valores mínimos y máximos
- Calcular la media y las desviaciones estándar
- Examinar la distribución de los datos.

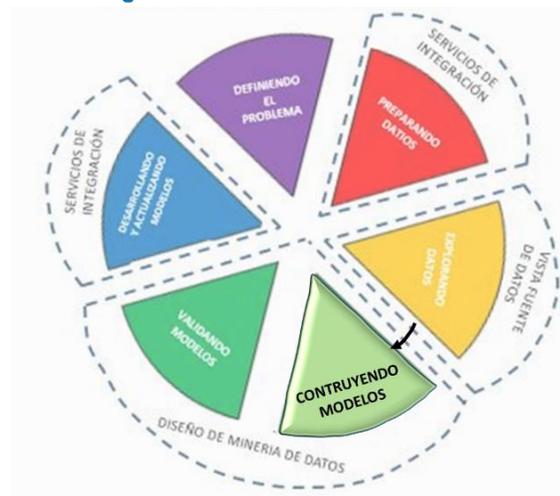
Por ejemplo, al revisar el máximo, el mínimo y los valores de la media se podrían determinar que los datos no son representativos de los clientes o procesos de negocio, y que por consiguiente debe obtener más datos equilibrados o revisar las suposiciones que son la base de sus expectativas. Las desviaciones estándar y otros valores de distribución pueden proporcionar información útil sobre la estabilidad y exactitud de los resultados. Una desviación estándar grande puede indicar que agregar más datos podría ayudarle a mejorar el modelo. Los datos que se desvían mucho de una distribución estándar se podrían sesgar o podrían representar una imagen precisa de un problema de la vida real, pero dificultar el ajustar un modelo a los datos.

Al explorar los datos para conocer el problema empresarial, puede decidir si el conjunto de datos contiene datos defectuosos y, a continuación, puede inventar una estrategia para corregir los problemas u obtener una descripción más profunda de los comportamientos que son típicos de su negocio.

3.3.4 Generar Modelo

El cuarto paso del proceso de minería de datos, como se exhibe en la figura 3.6, consiste en generar el modelo o modelos de minería de datos. Usará los conocimientos adquiridos en el paso Explorar los datos para definir y crear los modelos.

Figura 3.6 GENERAR MODELO



Deberá definir qué columnas de datos se utilizarán; para ello, creará una estructura de minería de datos. La estructura de minería de datos se vincula al origen de datos, pero en realidad no contiene ningún dato hasta que se procesa. Al procesar la estructura de minería de datos, se generan agregados y otra información estadística que se puede usar para el análisis. Cualquier modelo de minería de datos que esté basado en la estructura puede utilizar esta información.

Antes de procesar la estructura y el modelo, un modelo de minería de datos simplemente es un contenedor que especifica las columnas que se usan para la entrada, el atributo que está prediciendo y parámetros que indican al algoritmo cómo procesar los datos. El procesamiento de un modelo a menudo se denomina entrenamiento. El entrenamiento hace referencia al proceso de aplicar un algoritmo matemático concreto a los datos de la estructura para extraer patrones. Los patrones que encuentre en el proceso de entrenamiento dependerán de la selección de los datos de entrenamiento, el algoritmo que elija y cómo se haya configurado el algoritmo.

También puede utilizar los parámetros para ajustar cada algoritmo y puede aplicar filtros a los datos de entrenamiento para utilizar un subconjunto de los datos, creando resultados diferentes. Después de pasar los datos a través del modelo, el objeto de modelo de minería de datos contiene los resúmenes y modelos que se pueden consultar o utilizar para la predicción.

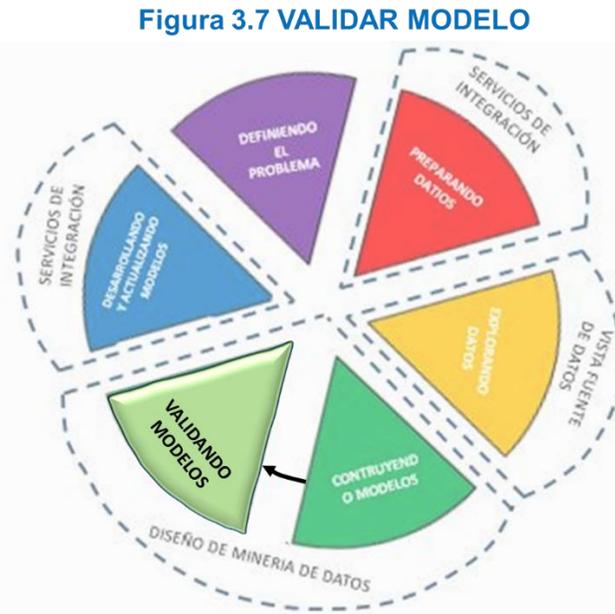
Es importante recordar que siempre que los datos cambian, debe actualizar la estructura y el modelo de minería de datos. Al actualizar una estructura de minería de datos volviéndola a procesar, se recuperan los datos del origen, incluido cualquier dato nuevo si el origen se actualiza dinámicamente, y vuelve a rellenar la estructura de minería de datos. Si tiene modelos que están basados en la estructura, puede elegir actualizar estos, lo que significa que se vuelven a entrenar con los nuevos datos, o pueden dejar los modelos tal cual.

A continuación, se muestra algunas de las técnicas de modelado:

- Redes bayesianas.
- Árboles de decisiones.
- Potenciación de gradientes.
- Redes neuronales.
- Bosques aleatorios.
- Máquinas de vectores de soporte.
- Modelos lineales generalizados
- Regresión Logística

3.3.5 Explorar y validar los modelos

El quinto paso del proceso de minería de datos, como se presenta en la figura 3.7, consiste en explorar los modelos de minería de datos que ha generado y comprobar su eficacia.



Antes de implementar un modelo en un entorno de producción, es aconsejable probar si funciona correctamente. Además, al generar un modelo, normalmente se crean varios con configuraciones diferentes y se prueban todos para ver cuál ofrece los mejores resultados para su problema y sus datos.

El conjunto de datos de entrenamiento se utiliza para generar el modelo y el conjunto de datos de prueba para comprobar la precisión del modelo mediante la creación de consultas de predicción.

Puede explorar las tendencias y patrones que los algoritmos detectan mediante los visores del diseñador de minería de datos. También puede comprobar si los modelos crean predicciones correctamente mediante herramientas del diseñador como el gráfico de mejora respecto al modelo predictivo y la matriz de clasificación. Para comprobar si el modelo es específico de sus datos o se puede utilizar para realizar inferencias en la población general, puede utilizar la técnica estadística denominada validación cruzada para crear automáticamente subconjuntos de los datos y probar el modelo con cada uno.

Si ninguno de los modelos que ha creado en el paso Generar modelos funciona correctamente, puede que deba volver a un paso anterior del proceso y volver a definir el problema o volver a investigar los datos del conjunto de datos original.

Algunos métodos utilizados en la validación de modelos se enlistan a continuación:

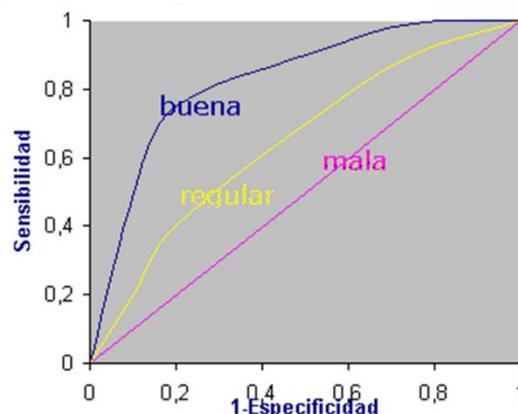
- Un *gráfico de elevación* (Lift) representa gráficamente la mejora que proporciona un modelo de minería de datos cuando se compara con una estimación aleatoria y mide el cambio en términos de una puntuación de *elevación*. Al comparar los puntajes de elevación para diferentes modelos, puede determinar qué modelo es el mejor.
- Error cuadrático medio (Yael, s.f.): Es una forma de evaluar la diferencia entre un estimador y el valor real de la cantidad que se quiere calcular, Mide el promedio del cuadrado del error, siendo el error el valor en la que el estimador difiere de la cantidad a ser estimada. Su fórmula es la siguiente:

$$ECM = \frac{1}{n} \sum_{i=1}^n (Y_i^{\wedge} - Y_i)^2$$

En el modelado estadístico, es usado para determinar la medida en la que el modelo no se ajusta a la información, o si el quitar ciertos términos puede simplificar el modelo de maneras benéficas. Proporciona una forma para elegir el mejor clasificador. El ECM coloca más peso en los errores grandes que en los pequeños (como resultado de elevar al cuadrado cada término), y por lo tanto enfatiza datos atípicos de maneras inconsistentes con la medida de los datos de la muestra.

- La curva ROC (Parra, 2017) es una representación gráfica del rendimiento del clasificador que muestra la distribución de las fracciones de verdaderos positivos y de falsos positivos. La fracción de verdaderos positivos se conoce como sensibilidad, sería la probabilidad de clasificar correctamente a un individuo cuyo estado real sea definido como positivo. La especificidad es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea clasificado como negativo. Esto es igual a restar uno de la fracción de falsos positivos. Es un método para evaluar clasificadores alternativos a la métrica expuesta, ROC (Receiver Operating Characteristic).

Figura 3.8 TIPOS DE CURVAS ROC

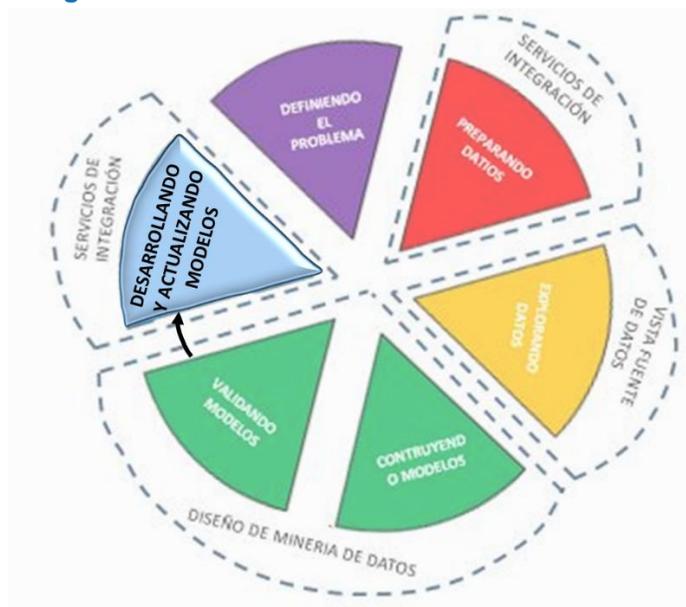


La curva ROC también es conocida como la representación de sensibilidad frente a (1-especificidad). Cada resultado de predicción representa un punto en el espacio ROC. El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo). Una clasificación totalmente aleatoria daría un punto a lo largo de la línea diagonal, que se llama también línea de no-discriminación. En definitiva, se considera un modelo inútil, cuando la curva ROC recorre la diagonal positiva del gráfico. En tanto que, en un test perfecto, la curva ROC recorre los bordes izquierdo y superior del gráfico. La curva ROC permite comparar modelos a través del área bajo su curva.

3.3.6 Implementar y actualizar los modelos

El último paso del proceso de minería de datos, como se señala en la figura 3.8, consiste en implementar los modelos que funcionan mejor en un entorno de producción.

Figura 3.9 IMPLEMENTAR Y ACTUALIZAR MODELO



Una vez que los modelos de minería de datos se encuentran en el entorno de producción, puede llevar a cabo diferentes tareas, dependiendo de sus necesidades. Las siguientes son algunas de las tareas que puede realizar:

- Use los modelos para crear predicciones que luego podrá usar para tomar decisiones comerciales.
- Crear consultas de contenido para recuperar estadísticas, reglas o fórmulas del modelo.
- Incrustar la funcionalidad de minería de datos directamente en una aplicación. Puede incluir Objetos de administración de análisis (AMO), que contiene un conjunto de objetos que la aplicación pueda utilizar para crear, cambiar, procesar y eliminar estructuras y modelos de minería de datos.
- Crear un paquete en el que se utilice un modelo de minería de datos para dividir de forma inteligente los datos entrantes en varias tablas. Por ejemplo, si una base de datos se actualiza continuamente con clientes potenciales, puede utilizar un modelo de minería de datos junto con Integration Services para dividir los datos entrantes en clientes que probablemente compren un producto y clientes que probablemente no compren un producto.
- Crear un informe que permita a los usuarios realizar consultas directamente en un modelo de minería de datos existente.

Actualizar los modelos después de la revisión y análisis. Cualquier actualización requiere que vuelva a procesar los modelos.

En el siguiente capítulo se describe el desarrollo del proyecto principal.

Capítulo 4 – Modelo de Minería

Durante mi trayectoria laboral uno de los proyectos más importante fue la creación de un Modelo de Minería, el cual he elegido como proyecto principal, los pasos desarrollados los describiré en este capítulo.

Es importante señalar que uno de los objetivos del área comercial fue generar conocimiento en sus empleados para la generación de modelos de minería de manera interna y no solo utilizar a proveedores externos para este fin. Para lograr lo anterior la empresa envió a varios de sus empleados a un diplomado de Minería de datos que fue impartido en el ITAM (Instituto Tecnológico Autónomo de México). Lograr uno de estos lugares significada tener conocimientos en Base de Datos (Análisis de información), Algoritmos y Estadística.

Después de ser postulada por mi director y obtener un lugar aprobando el curso propedéutico, asistí a dicho diplomado, el cual me mostró una visión diferente de como ver los datos que hasta el momento no tenía. En este proyecto no solo tuve el rol de líder de sistemas, sino también participe como BP y usuario.

Como líder de sistemas conté con un consultor Oracle, y yo fungí como consultor SAS, la duración es este proyecto fue de 11 meses, realizando la liberación a producción.

4.1 Situación actual

La Empresa como negocio de banca múltiple y como líder en productos financieros ha enfocado sus esfuerzos en mantener dicho liderazgo, haciendo hincapié en los diferentes indicadores que le proporcionen una mejor calidad de la información de cada una de sus bancas.

Actualmente la banca de empresas y gobierno tiene un margen considerable de utilidad para el negocio, por lo anterior, el banco se ha dado a la tarea de investigar la efectividad de sus estrategias y campañas para este foco de clientes, con el objetivo de ser más asertivos y lograr un crecimiento mayor al actual.

Partiendo de que la estrategia más importante es la Captación de Saldos como uno de los principales generadores de valor al banco se ha desarrollado este proyecto.

4.2 Objetivo

Generar un modelo de Minería que identifique aquellas Empresas de la Banca que cuentan con el producto de captación y tiendan a incrementar el saldo de captación (producto Vista Moneda Nacional) de manera relevante. Utilizando la técnica de Up Selling (Electronico, s.f.) localizar a los clientes con un incremento de saldo notorio para desarrollar el interés en la compra de un producto de mayor calidad y obtener un mayor beneficio para la empresa.

4.3 Desarrollo de la solución

Antes de explicar la solución es importante mencionar dos cosas, la primera es que para concluir con el diplomado de Minería de Datos el proyecto final consistió en desarrollar un modelo con el objetivo de implementarlo dentro de la empresa, al pertenecer a un área de sistemas y no de negocio, realicé equipo con un usuario del área de Empresas y Gobierno (la cual contaba con información a nivel negocio para este desarrollo), la segunda cuestión es que anteriormente yo participe en la implantación de modelos de minería, modelos ya desarrollados (por empresas externas), es decir, una vez desarrollado el modelo, nos enfocábamos en automatizar la entrada, el modelo y su respectivas salidas, pero no en la generación del modelo en sí.

Al tratar de integrar la metodología corporativa con la metodología de Modelos de Minería, para generar un plan de trabajo, detecté que no podría definir las iteraciones necesarias en la generación de modelo, ni el tiempo que me llevaría esto, dada mi inexperiencia en la generación de modelos y que mi compañero de equipo y yo debíamos trabajar en nuestras actividades diarias, por lo que el tiempo disponible para esta tarea fue reducido. Tomé la decisión de dividir este proyecto en las siguientes etapas:

- Etapa 1 - Creación del modelo.
- Etapa 2 – Implantación del modelo.

En la figura 4.1 se muestra gráficamente el detalle de las etapas:

Figura 4.1 Etapas para el Modelo de Minería



De acuerdo al diagrama, la Etapa 1 - Creación del modelo, realizaríamos los pasos de un Modelo de Minería, en esta etapa participamos mi compañero del diplomado y yo. Después en la segunda etapa - Implantación del Modelo, utilice las fases de desarrollo de software de la metodología corporativa y me apoye en mi equipo de trabajo para la realización de algunas tareas.

4.3.1 Creación del Modelo

A continuación, se describen los pasos realizados para la etapa 1 correspondiente al ciclo de un Modelo de Minería.

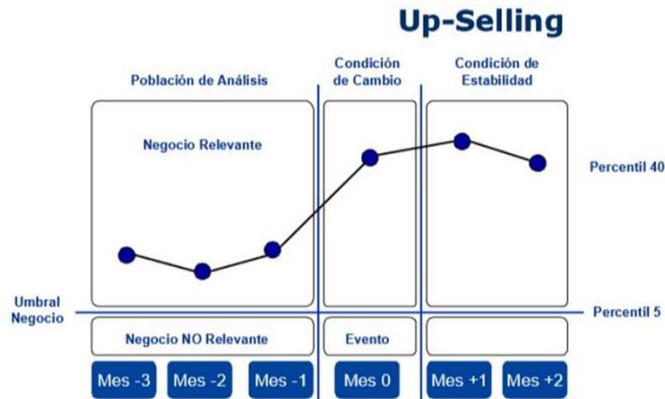
4.3.1.1 Paso 1-Definir el Problema

El primero fue definir el problema, este problema básicamente es el objetivo del proyecto: Identificar a las Empresas de la Banca de Empresas y Gobierno que cuentan con productos de captación y tiendan a incrementar el saldo de captación de manera relevante.

La variable target de los modelos de Up selling lo definió mi compañero de equipo como un incremento relevante en el monto o en el número de operaciones en el producto de interés. Dicho incremento tiene que ser anticipado por el cumplimiento de unas condiciones de estabilidad ex ante y seguido por el cumplimiento de unas condiciones de estabilidad ex post.

En la figura 4.2 se muestra gráficamente el comportamiento buscado.

Figura 4.2 Up-Selling



Una vez definido el problema nos enfocamos a identificar el tipo de información que maneja el usuario de negocio (este punto lo definió mi compañero de equipo dado que él conocía dicha información). En la figura 4.3 se muestra lo que encontramos:

Figura 4.3 INFORMACION PARA EL MODELO

SALDOS	CAPTACION Y CARTERA
MONTOS Y OPERACIONES EN INF	COBRANZA, DISPERSION Y TRANSACCIONALIDAD
RENTABILIDAD	IF, INF, COSTOS, BAI TOTAL
PERFIL DEL CLIENTE	SECTOR, VENTAS Y UBICACION
VINCULACION	PRODUCTOS DEL CLIENTE

La información se encontraba almacenada en una base de datos productiva del área, lo cual nos facilitó bastante el trabajo:

- Perfil - CTE_PERFIL
- Monto - CTE_MONTO
- Cartera - CTE_CARTERA

4.3.1.2 Paso 2-Preparar Datos

Una vez definido lo anterior pasamos al paso 2 que fue Preparar los Datos, nos encontramos la mayor parte de la información en un ambiente productivo, sin embargo, no contaba con suficiente historia para generar el modelo, por lo que tuvimos que recurrir al servidor local del área de Inteligencia comercial de Empresas y Gobierno, quienes tenían almacenados los datos en archivos planos, así como en archivos tipo Excel.

El trabajo continuó al extraer esta información, tratar de completar la historia necesaria que fue de dos años, estas actividades las desarrollé personalmente, todo esto se realizó previo a la creación de la tabla de entrenamiento del modelo.

La siguiente tarea fue revisar la población de análisis y nos enfocamos en el sobre muestreo para tener datos suficientes, considerando lo siguiente:

- Se permite sólo el paso a clientes que sean población de análisis del modelo.
- En caso de que la tasa de eventos fuera menor a un 10% se realiza un sobre muestreo.

La técnica de sobremuestreo empleada consistió en tomar parte de los casos en que no haya evento, de tal manera que en la tabla resultante se obtenga un 10% de “unos” en la variable tarjet (variable indicadora de evento de éxito). El filtro de las observaciones que no dan lugar a eventos opera de forma aleatoria sin generar distorsiones en el conjunto de entrenamiento.

Para realizar la limpieza de los datos se revisaron los siguientes aspectos:

- Revisar porcentaje de nulos, rango de variables: esto ayuda a encontrar errores en el cálculo de algunas variables.
- Tipos de datos: nominal, ordinal, intervalo... (por ejemplo: “CD_POSTAL_CTE” es Nominal).
- Establecer como “target” todas aquellas variables que vayan a ser indicador de evento de éxito en algún modelo.
- Rechazarlas variables que no se desea incluir en la modelización.

Se generaron tablas de usuario en la base de datos con los filtros correspondientes en cada variable, esta tarea la realice por medio de queries (sql) y los filtros aplicados fueron definidos en su mayoría por mi compañero de equipo.

4.3.1.3 Paso 3 – Exploración de Datos

El paso 3 exploración de datos nos llevó más tiempo, por mi parte no conocía los conceptos que esta área de negocio manejaba, por lo que mi compañero de equipo y yo tuvimos varias reuniones donde me explicó dichos conceptos.

Realizamos una primera exploración de variables explicativas identificando los perfiles que caracterizan las poblaciones con mayor propensión al éxito. Este análisis exploratorio nos permitió guiar los pasos futuros y nos dio una primera caracterización del público objetivo.

Para cada modelo, las estadísticas descriptivas se realizaron de la siguiente manera:

- Discretización (generalmente por percentiles) de la población de análisis.
- Estudio del ratio de éxito para cada categoría

En esta etapa también exploramos posibles efectos de variables complejas, determinadas desde un punto de vista de negocio y que pueden no estar incluidas en el conjunto de variables explicativas inicial.

Observamos que el comportamiento de las variables de manera general fue la esperada, sin embargo, también se detectó de acuerdo a las estadísticas obtenidas que la cantidad de valores missing era considerable por lo que se llegó a la decisión de sustituir con 0 estos valores y poder analizar mejor los datos.

Para el análisis multivariante que realizamos a la información se utilizó la técnica de componentes principales para obtener la matriz de correlación, así detectamos las variables correlacionadas y procedimos a eliminarlas como entrada al modelo.

A continuación, se detallan el tratamiento que realizamos a la tabla de entrenamiento, en base al análisis estadístico determinamos las siguientes acciones sobre los datos:

- Descarte de algunas variables.
- Agrupación de variables para disminuir el número de las mismas.
- Imputación de los valores missing.
- Transformación de variables con valores extremos para su normalización.

En la figura 4.4 se muestran los nodos utilizados para la generación de la población de análisis.

Figura 4.4 NODOS DEL MODELO



Una vez realizado lo anterior procedimos a la generación del tablón de entrenamiento, esta información se convirtió así en nuestra **población de análisis**.

Considerando que nuestro evento se trata de un up-selling entonces la población de análisis estuvo conformada por lo siguiente:

- Aquellas empresas que poseen un producto de captación.
- Aquellas empresas que en un principio no lo poseían y después lo contrataron.
- No considerar a clientes con saldos menores a treinta mil durante los últimos 6 meses.

4.3.1.4 Paso 4 – Construcción de Modelos

Lo siguiente que realizamos fue el paso 4 crear modelos, la herramienta utilizada para la generación del modelo de minería fue SAS Miner, dado que el tipo de modelo a desarrollar fue un modelo predictivo se decidió aplicar los modelos:

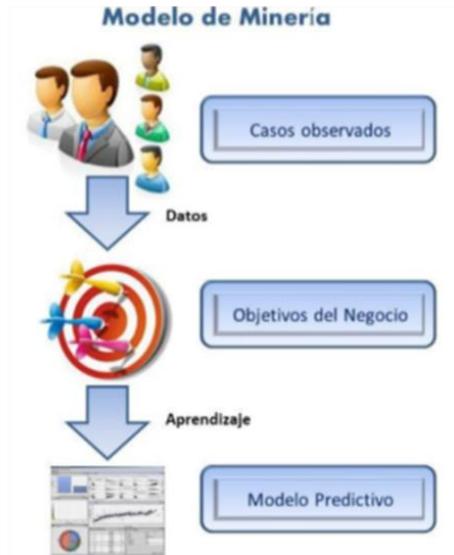
- Los árboles de decisión.
- Las redes neuronales.
- Regresión Lineal.

La construcción de un modelo se realiza a través de dos pasos, uno inicial de aprendizaje en base a los patrones de comportamiento de los clientes que extraemos a partir de los datos recogidos en transacciones, tenencia de productos e interacciones con el cliente.

Así, el modelo se transforma en un programa que, a partir de los datos observados para un cliente, es capaz de predecir cuál será el comportamiento del mismo para el evento que tiene asociado.

En la figura 4.5 se plasma gráficamente lo anterior.

Figura 4.5 PASOS DE UN MODELO



De acuerdo a las estadísticas obtenidas de la variable objetivo consideramos necesaria una sobrecarga en la tabla de entrenamiento y comparamos los modelos aplicados.

4.3.1.5 Paso 5 – Validación de Modelos

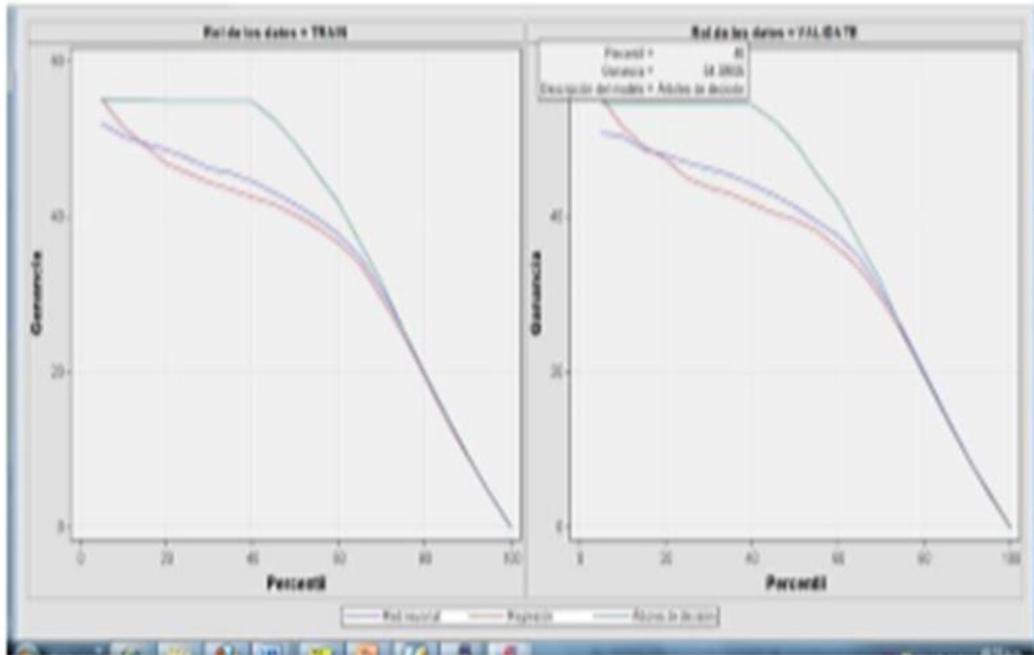
El paso 5 fue la validación de modelo, este proceso se desarrolló con un conjunto de datos que era independiente a los datos utilizados para validar el modelo. Que nos permitió así poder valorar la capacidad predictiva de manera objetiva. Utilizamos solo algunas técnicas de validación de Modelos, recomendadas por el profesor que nos impartió el Diplomado de Minería.

Realizamos la estimación del conjunto de validación, que recogió el 40% de los datos iniciales, y se construyó el gráfico LIFT (medida de efectividad de un modelo predictivo). Se calcula como la razón de los resultados obtenidos con y sin el modelo.

La curva LIFT midió la ganancia en la respuesta con modelo con relación a lo que sucedería sin el modelo. Si observamos el siguiente grafico de Lift el percentil 40 muestra un valor de 54.39925, interpretamos que en el 40% más propenso de la población (según cuanto estimado por el modelo) la tasa de eventos reales observada es 54.39925 para el árbol de decisión.

En la Figura 4.6 se muestra la curva Lift obtenida durante la generación del modelo.

Figura 4.6 Grafico Lift



El gráfico de porcentaje de “respuesta capturada” que generamos corresponde a la distinta tasa de evento que tienen los clientes al ordenarlos según la estimación de propensión realizada por el modelo.

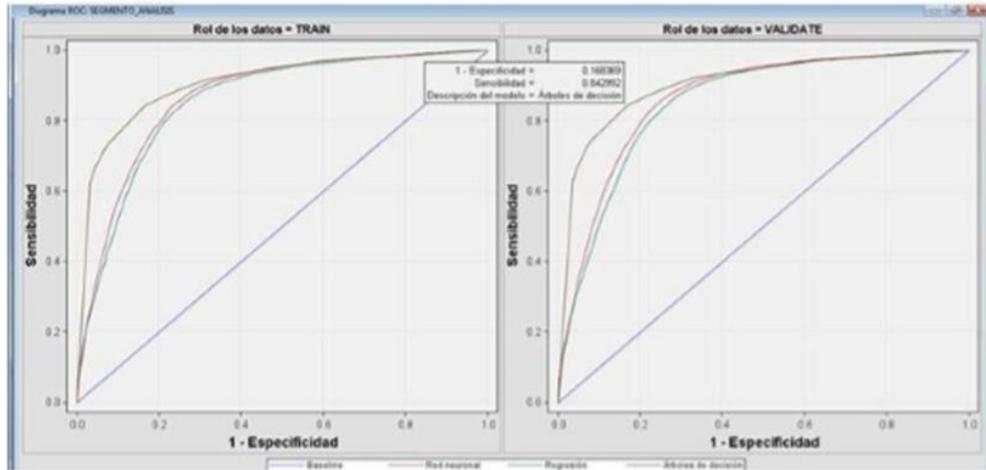
Por ejemplo: el valor que la función toma en el número 40 del eje de las abscisas (percentiles) corresponde a la tasa de evento real observada en el 40% de las empresas que el modelo considera las más propensas a dar lugar al evento para el árbol de decisión.

En la tabla de Comparación de estadísticos de los 3 modelos pudimos observar que el error cuadrado promedio más bajo corresponde al árbol de decisión con 0.115423. De acuerdo a la matriz de clasificación observamos que el modelo que menos se equivoca en predecir la variable objetivo es la regresión.

Se graficó la sensibilidad contra 1 - Especificidad, ROC, y sabíamos que mientras más se aleje ROC de la curva de 45 grados hacia la esquina superior izquierda, el modelo es mejor. Es decir, acierta tanto en positivos como en negativos. Pudimos observar que 1- Especificidad es igual a 0.168369 y la sensibilidad es igual a 0.842992 en los datos de validación para el árbol de decisión.

En la Figura 4.7 se muestra la curva ROC obtenida durante la validación del modelo.

Figura 4.7 Curva ROC



Una vez que decidimos las variables que entrarán a explicar el modelo y las tramificaciones definitivas de las mismas, procedimos al cálculo del modelo final seleccionado de forma directa las variables elegidas. Éste será el resultado de calcular los parámetros de las variables sobre todo el conjunto de entrenamiento inicial. Con los resultados anteriores determinamos que el modelo a implementar sería el árbol de decisiones.

4.3.1.6 Paso 6 – Implementación del Modelo

Una vez que seleccionamos el modelo se incorporó al diagrama el nodo “score” que calcula el código de scoreo. El modelo se transforma en un programa (código fuente) que se utiliza para predecir la probabilidad del evento de interés en el conjunto de los datos de explotación. Especificamos los tipos de datos, de forma que la tabla de salida sea una tabla tipo SAS perfectamente exportable al entorno de desarrollo.

Decidí solicitar apoyo al área de Gestión de iniciativas, Desarrollo y Gestión del servicio, para que nuestro modelo desarrollado en el Diplomado de Minería pudiera quedar implementado productivamente, dado que este proyecto no se encontraba dentro del plan anual de proyectos, los cuales tiene línea presupuestal asignada, realice las negociaciones necesarias para el apoyo de las diferentes áreas.

A continuación, se describe la segunda etapa del proyecto: Implantación del Modelo de Minería, esta etapa abarca las fases del diseño y desarrollo (cap. 2.1).

4.3.2 Implantación del Modelo de Minería

De acuerdo a la metodología corporativa el usuario del área Inteligencia Comercial de la Banca de Empresas y Gobierno debería levantar un requerimiento a través de un A825– Solicitud de requerimiento ante su Socio de Negocio, quien se encargó de recabar la información necesaria para profundizar más sobre la necesidad del usuario y genera el C100 – Diseño de la Solución, el cual debe ser entregado al área de sistemas para este caso a Analítica Central, para que diera su respuesta en el formato C102 – Alcance funcional, donde se especificaron las actividades a realizar, así como los tiempos y costos que implicaba este desarrollo. Las anteriores actividades fueron generadas por mí, dado que yo participe en la generación del modelo y conocía todas las características del requerimiento a solicitar, por lo cual generé los documentos.

Una vez generado el modelo, procedí a su liberación en producción. A continuación, se detallan las fases del Desarrollo.

4.3.2.1 Fase de Análisis

En la generación del formato C-102 determiné el tiempo del plan del proyecto que sería de 5 meses, más 4 meses del periodo de garantía debido a que su ejecución sería bimestral, decidí asignar a 2 personas para el Desarrollo de este proyecto, un consultor Oracle y un consultor SAS, el rol de consultor SAS fue desarrollado por mí.

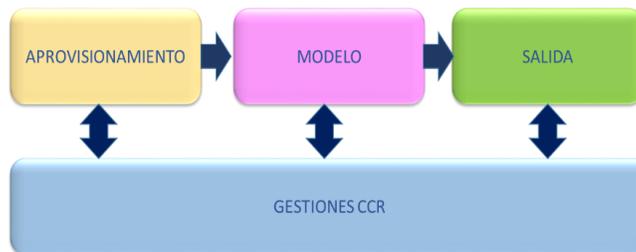
Durante la fase de Análisis se revisaron los requisitos solicitados por el usuario para crear las características, funciones y funcionalidades del sistema, para lo anterior realice reuniones con el consultor Oracle, donde analizamos el requerimiento, dada mi experiencia como líder de proyecto y en específico con proyectos de modelos de minería esta fase no requirió mucho tiempo, dado que en el documento C100 no solo se especificaron las consultas y código del modelo, también se hizo entrega de los programas generados durante la creación del modelo, con estos objetos las actividades correspondientes al consultor SAS y al consultor Oracle fueron distribuidas de manera más eficiente, se clasificaron las tareas de la siguiente manera:

- **Aprovisionamiento** – 322 campos a aprovisionar, la información se generaría de forma bimestral, la fuente de información sería la base de datos del área de inteligencia comercial de Empresas y Gobierno, así como los nombres de las tablas a utilizar para dicho aprovisionamiento, los cruces y las variables calculadas.

- Modelo – Identificaron las rutas de las librerías, fechas, fuente de información, código del modelo, entre otros.
- Salida – Se identificó la tabla de salidas, así como la fuente destino de almacenamiento.
- Gestiones – Se identificaron las gestiones necesarias para el proyecto como: solicitud de espacio, permisos, creación de tablas y puesta en producción.

En la figura 4.8 se muestra gráficamente la división de tareas descritas anteriormente de este desarrollo.

Figura 4.8 FASE DE ANALISIS



4.3.2.2 Fase de Diseño

En la fase de Diseño mi equipo y yo generamos los prototipos con las especificaciones productivas de acuerdo con las políticas correspondientes a cada uno de los ambientes a utilizar en esta implantación, en este caso se utilizaron dos ambientes distribuidos.

El primer ambiente corresponde a la Base de Datos, donde se almacenarían los queries para consultas y transformaciones correspondientes en la información que fungió como entrada al modelo, las características en las estructuras de datos de las tablas generadas y la respectiva comunicación entre los ambientes de este desarrollo. El segundo ambiente fue el servidor SAS, donde se ejecutaría el modelo, en este ambiente se identificaron las rutas productivas y las interfaces de salida.

En la figura 4.9 se describe gráficamente los ambientes y cada uno de los componentes generales:

Figura 4.9 FASE DE DISEÑO



Así también se diseñaron los planes de las pruebas funcionales y de usuario, se detallaron los casos de prueba que se enviaron al usuario para su aceptación.

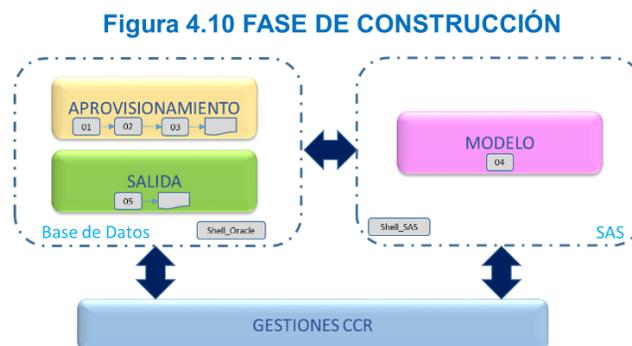
Por último, se diseñó la cadena de procesos a ejecutar, con sus especificaciones, orden de ejecución de cada componente y códigos de salida autorizados por el área de Producción Distribuida para continuar con el proceso, lo anterior con el fin de iniciar a tiempo la gestión con el área de Producción Distribuida para la calendarización correcta del sistema. Dado que tenía experiencia en la puesta en producción de modelos de minería previos, esta fase no requirió mucho tiempo.

4.3.2.3 Fase de Construcción

En la fase de Construcción se realizó la codificación de cada uno de los programas establecidos en la fase de Diseño, a continuación, se enlistan de manera general los programas más relevantes del desarrollo:

- 01_generacion_tablas.sql: Este archivo contiene los códigos que unifica las tablas de entrada.
- 02_agrupación_cte.sql: Este archivo contiene la agrupación de los contratos a nivel cliente.
- 03_envio_tabla.sql: Este archivo contiene el envío de la tabla generada para el modelo a una tabla con formato SAS.
- 04_modelo_upselling.sas: Este archivo contiene el código del modelo, parametrizado en sus entradas y salidas.
- 05_salida_modelo.sas: Este archivo contiene el código de carga para la tabla de salida del modelo.
- Shell_Oracle.sh: Este archivo contiene el Shell diseñado para la ejecución del proceso Oracle.
- Shell_SAS.sh: Este archivo contiene el Shell diseñado para la ejecución del modelo y su salida.

En la figura 4.10 se visualizan los componentes descritos anteriormente dentro de los ambientes correspondientes:



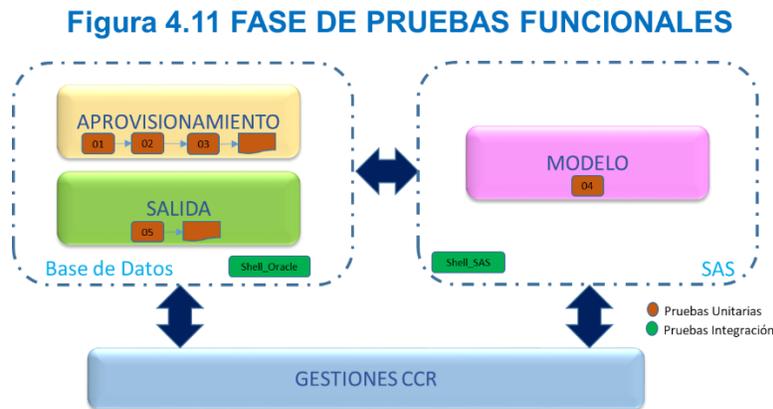
En esta fase también se levantó el requerimiento al área de Base de Datos para la creación de las tablas productivas para la entrada y salida del modelo, una vez generadas se solicitó la creación de los objetos necesarios en el ambiente de pruebas para la siguiente fase del desarrollo, así también se determinó la muestra de registros a utilizar para la fase de pruebas, esta muestra fue entregada por el usuario.

4.3.2.4 Fase de Pruebas Funcionales

En la fase de Pruebas Funcionales, ejecutamos primero las pruebas unitarias para determinar posibles fallas en cada uno de los componentes, se detectaron algunos errores, que fueron corregidos de manera inmediata, y así evitamos la afectación futura a la funcionalidad del sistema.

Una vez que terminamos con las pruebas unitarias, continuamos con las pruebas de integración que nos permitieron comprobar el funcionamiento de forma integral de todos los componentes del sistema, en estas pruebas no encontramos incidencias.

En la figura 4.11 se muestra gráficamente las pruebas realizadas al sistema.



Una vez concluidas las pruebas funcionales, se inició con la gestión del pase a producción con el CCR, específicamente con el área de Control-M, con el fin de calendarizar la automatización del proceso.

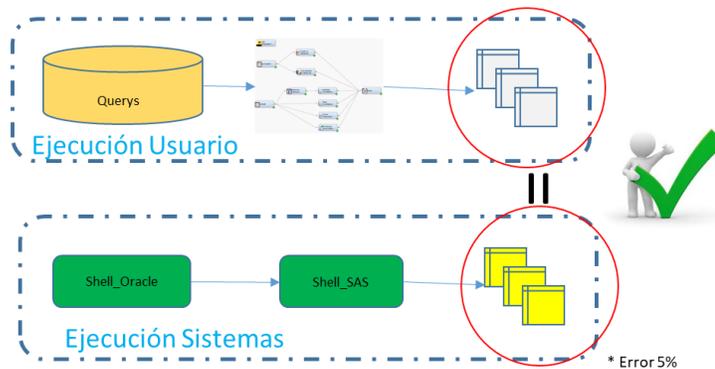
4.3.2.5 Fase de Pruebas de Usuario

En la fase de Pruebas de Usuario, el usuario realizó sus respectivas pruebas, las cuales fueron especificadas en el plan de pruebas de usuario establecido en la fase de diseño y con la muestra de información entregada por él.

Estas pruebas consistieron en que mi equipo y yo realizaríamos la ejecución del modelo en el ambiente de pruebas y exportaríamos las tablas de salida del modelo a la base de datos, pero en el ambiente de trabajo del usuario, mientras que él usuario ejecutaría el modelo en su ambiente de trabajo, es decir Sistemas ejecutaría el modelo mediante shell's y el usuario ejecuto el modelo por medio de querys en la Base de Datos (tabla de entrada) y código en el Enterprise Miner (modelo), una vez obtenidas las respectivas tablas de salida, éstas fueron comparadas por el usuario, con un criterio de aceptación del 5% de diferencia entre resultados. El sistema fue ejecutado 3 veces con ventanas de tiempo diferentes.

En la figura 4.12 se presenta gráficamente el proceso que se llevó acabo en esta fase.

Figura 4.12 FASE DE PRUEBAS DE USUARIO



Una vez concluidas las pruebas de usuario se procedió a la reunión para la aprobación del desarrollo, el usuario firmó el formato A990 – Aprobación de usuario.

4.3.2.6 Fase de Implantación

Una vez concluidas las pruebas de usuario, y con firmado el formato A990 se continuó con la fase de implementación, realizamos las gestiones para el pase a producción de los componentes, su debida calendarización, así como su supervisión en el periodo de garantía de sistema. Es decir, el formato A990 fue entregado a la Oficina de Proyectos y también al área de Control-M para la liberación de los componentes a Producción.

Una vez realizada la liberación en Control-M procedimos a esperar las 2 primeras ejecuciones productivas, éstas ejecuciones no presentaron ninguna falla y con esto se nos otorgó la aprobación del área de Producción para cerrar el plan de trabajo con la Oficina de Proyectos y con esto se dio por concluido el desarrollo.

En la figura 4.13 se muestran los componentes a implantar de manera general que se llevó a cabo en esta fase.

Figura 4.13 FASE DE IMPLANTACIÓN



Los espacios necesarios para el desarrollo de este proceso fueron los siguientes:

- 5 GB para los datos iniciales.
- 15 GB para el espacio de trabajo.
- 2 GB para la salida.

El tiempo aproximado que nos llevó esta etapa fue de 4 meses.

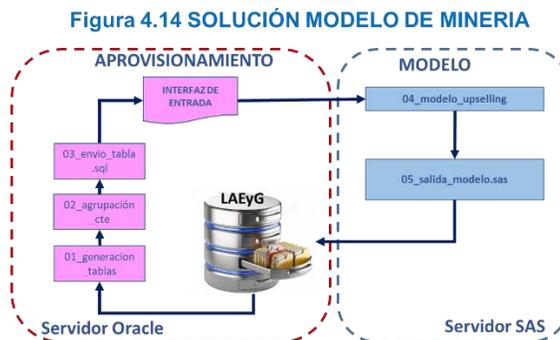
4.5 Resultados

Con la información generada por el modelo el usuario del Negocio logra:

- La identificación de perfiles potenciales para el incremento de su saldo.
- Generación oportuna de la siguiente mejor acción comercial.
- Optimización de esfuerzos y recursos.
- Obtención de resultados a corto plazo.

Una vez que el modelo genera el porcentaje de probabilidad de incremento de saldo en el producto Vista Moneda Nacional, se seleccionaba a aquellos clientes que estaban por arriba del 85% de probabilidad, esta información se enviaba al área de Inteligencia comercial, quien dictaminada la mejor acción comercial para cada uno de los prospectos.

En la figura 4.14 se muestra la solución implementada para este proyecto.



Capítulo 5 - Conclusiones

La realización del presente trabajo ha sido muy significativa para identificar y resaltar los puntos relevantes en mi trayectoria profesional; Considero que mi educación en la carrera de Ingeniería en Computación me ha dado las bases para el aprendizaje y el desarrollo de mis habilidades de análisis, diseño, planificación y control de los sistemas computacionales, así como la gestión de recursos materiales y humanos, entre otros.

Considero que los proyectos expuestos en este trabajo han dejado una huella en mi trayectoria profesional por las diferentes tecnologías utilizadas, la rápida adaptación a la metodología corporativa, el trabajo y la confianza en mi equipo de trabajo, el control de riesgos, la solución de conflictos, entre otros, que de una u otra manera se convirtieron en un reto.

El desarrollo del proyecto de Minería para la detección de empresas con potencial de crecimiento en saldo de Captación lo considero el desarrollo más importante hasta el momento, a pesar de no estar dentro de los proyectos estratégicos de la empresa, lo anterior debido a que recibí de la empresa la oportunidad de asistir al “Diplomado en Minería de Datos” impartido en el ITAM, el cual fue diseñado para personal de La Empresa, al implementar este desarrollo considero que estoy pagando con creces la inversión que el Banco realizó en mí.

La mejor forma de continuar y avanzar en mi trayectoria profesional es adaptarme lo más rápido posible al cambio, ya que las soluciones tecnológicas se mueven de enfoques tradicionalistas a metodologías más ágiles, para entregas más rápidas y a menor costo sin olvidar la calidad en el desarrollo.

Al terminar los estudios de Ingeniería en Computación uno se convierte en un profesional, sin embargo, la experiencia es la que te convierte en un profesional.

Glosario

Ad-hoc: Los informes ad hoc se utilizan para generar informes que combinan de forma temporal distintas identidades como categoría de datos, período, tipo de consolidación, compañía, tipo de moneda y dimensión adicional. Sólo los informes basados en formularios de entrada de datos se pueden utilizar en informes ad hoc.

DWC: Data Warehouse Corporativo

Entrenamiento del modelo: Proceso mediante el cual se ajustan los parámetros del modelo.

Ex ante: Es una palabra neolatina que significa "antes del suceso".

Ex post: Es una locución latina que significa "después del hecho".

Fase: Grado de avance de un proyecto expresado según la metodología Tecnología y Operaciones (TyO).

Iniciativa: Es una idea, o parte de una idea, de la que se están concretando los requisitos y de la que se está evaluando la mejor alternativa de solución, es decir, una iniciativa es una idea sobre la que un BP ha empezado a trabajar. Una vez autorizada por el Negocio la iniciativa se convierte en uno o varios proyectos.

Población de análisis (Conjunto de “train”): población elegida para la construcción del modelo. Debe contener un número mínimo de clientes que han realizado el comportamiento buscado durante un cierto periodo de tiempo anterior al momento actual. Dichos clientes deben tener las mismas características que los clientes sobre los cuales se aplicará el modelo en el futuro. La calidad de esta selección inicial determinará la complejidad máxima del modelo

Proyecto: Es una iniciativa autorizada por el Negocio y que entra en el ciclo de desarrollo. Con el fin de disminuir la complejidad y dar respuesta al negocio, los proyectos deberían durar menos de un año.

Scoring: El modelo se utilizará para predecir en el futuro el conjunto de la población con mayor potencial o probabilidad (en cuanto a la variable estudiada).

Referencias:

- Duncan, O. (04 de 2018). *Microsoft*. Obtenido de Conceptos de minería de datos:
<https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017>
- Electronico, C. (s.f.). *Upselling – ¿qué es y cómo se hace?* Obtenido de
<http://www.marketingguerrilla.es/upselling-que-es-y-como-se-hace/>
- Parra, F. (15 de junio de 2017). *Estadística y Machine Learning con R*. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/293405_4029f1f23f834b7195189d5504a436b2.html
- Sanjuán, F. J. (s.f.). *economipedia*. Obtenido de Outlier:
<https://economipedia.com/definiciones/outlier.html>
- Sanjuán, F. J. (s.f.). *economipedia.com*. Obtenido de Outlier:
<https://economipedia.com/definiciones/outlier.html>
- SAS. (22 de 03 de 2019). Obtenido de Minería de datos:
https://www.sas.com/es_mx/insights/analytics/data-mining.html
- Yael, G. C. (s.f.). *prezi.com*. Obtenido de Minería de Datos:
<https://prezi.com/rz6jpfubplae/mineria-de-datos-matriz-de-confusion-media-geometrica-error-cuadratico-medio-varianza-muestral-desviacion-estandar/>