

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE
MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y
COMPUTACIÓN**



**SISTEMA DE GESTIÓN Y SOPORTE DE TOMA DE
DECISIONES BASADO EN ALGORITMOS DE
BAYES Y CLUSTER PARA MEJORAR LOS
PROCESOS ANALÍTICOS DEL ÁREA COMERCIAL
DE UNA EMPRESA EDUCATIVA**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

CARLOS ENRIQUE QUEPUY PERLECHE

Chiclayo 29 de Septiembre del 2016

**SISTEMA DE GESTIÓN Y SOPORTE DE TOMA DE
DECISIONES BASADO EN ALGORITMOS DE
BAYES Y CLUSTER PARA MEJORAR LOS
PROCESOS ANALÍTICOS DEL ÁREA COMERCIAL
DE UNA EMPRESA EDUCATIVA**

POR:

CARLOS ENRIQUE QUEPUY PERLECHE

**Presentada a la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR EL JURADO INTEGRADO POR

Ing. Héctor Miguel Zelada Valdivieso
PRESIDENTE

Ing. Huilder Juanito Mera Montenegro
SECRETARIO

Ing. Segundo José Castillo Zumarán
ASESOR

DEDICATORIA

*A mis padres Ricardo y Gladys, por ser mi estímulo para luchar sin déjame
vencer por los obstáculos.*

*A mis hermanos Ricardo y Fernando, por su apoyo y comprensión para
culminar este trabajo de investigación.*

AGRADECIMIENTO

A Dios, por ser mi Padre y guía.

A mi asesor Segundo José Castillo Zumarán por sus consejos oportunos y su generosa ayuda, a el mi profunda gratitud.

Al jurado de tesis por sus aportes en la culminación de la investigación.

ÍNDICE

DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
RESUMEN.....	8
ABSTRACT.....	9
I . INTRODUCCION.....	10
II. MARCO TEÓRICO.....	12
2.1 Antecedentes.....	12
2.2 Bases Teórico Científicas.....	15
2.2.1 Toma de Decisiones.....	15
2.2.2 Business Intelligence.....	18
2.2.3 Algoritmo de Bayes.....	21
2.2.4 Algoritmo de Clúster.....	22
2.2.5 Metodologías de sistemas de Inteligencia de Negocios.....	23
2.2.6 Aporte de las aplicaciones de Inteligencia de Negocios a la Tesis..	25
III. MATERIALES Y MÉTODOS.....	25
3.1 Diseño de investigación	25
3.2 Metodología.....	28
IV. RESULTADOS	
4.1 Planeamiento del proyecto.....	30
4.1.1 Identificación del escenario del Proyecto.....	30
4.1.2 Identificación de los Usuarios y áreas intervinientes.....	31
4.1.3 Análisis de Factibilidad.....	31
4.2 Definición de los Requerimientos del Negocio.....	33
4.2.1Requerimientos del Negocio.....	33
4.2.1.1 Requerimientos de usuarios finales.....	33
4.3 Modelo Dimensional.....	36
4.3.1 Análisis de Medidas y Dimensiones.....	36
4.3.2 Star Net para el análisis.....	36
4.3.3 Análisis de las Dimensiones y Jerarquías.....	38
4.4 Modelo de Datos Dimensional.....	40
4.5 Diseño Físico.....	41
4.6 Diseño y Desarrollo de Presentación de Datos.....	46
4.7 Diseño de la Arquitectura Técnica.....	50
4.8 Especificación de aplicaciones para los usuarios finales.....	51
4.9 Implementación.....	51
4.10 Mantenimiento.....	51
5. Desarrollo de aplicaciones para usuarios finales.....	52
V. DISCUSION.....	67
VI. CONCLUSIONES.....	80
VII. REFERENCIAS BIBLIOGRÁFICAS.....	81
VIII. ANEXOS.....	84

ÍNDICE DE GRÁFICOS.

Gráfico N° 01. Visor de Bayes Naives.....	22
Gráfico N° 02. Business Dimensional Lifecycle (BDL) Ralph Kimball.....	24
Gráfico N° 03. Organigrama de la empresa Eknowledge Group S.A.C.....	31
Gráfico N° 04. Star net ingreso.....	36
Gráfico N° 05. Star net demanda.....	37
Gráfico N° 06. Diseño de Base de Datos Dimensional.....	40
Gráfico N° 07. Poblado hechos ingreso y demanda.....	46
Gráfico N° 08. Arquitectura Técnica de Business Intelligence.....	50
Gráfico N° 09. Acceso a Usuario EknowledgeGroup SAC	52
Gráfico N° 10. Menú Principal del Sistema de Gestión y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.....	53
Gráfico N° 11. Dashboard Demanda.....	54
Gráfico N° 12. Reporte de los cursos con la cantidad de inscritos y retiros.....	55
Gráfico N° 13. Ocurrencias cliente por área.....	56
Gráfico N° 14. Dashboard Ingreso.....	57
Gráfico N° 15. Ciclo de vida de un producto.....	58
Gráfico N° 16. Evolución Ingreso.....	59
Gráfico N° 17 Ingresos Bruto por Producto.....	60
Gráfico N° 18 Evaluación Docente.....	61
Gráfico N° 19 Reporte Demanda e Ingreso.....	62
Gráfico N° 20 Elaboración de Clúster.....	63
Gráfico N° 21 Perfiles de Clúster.....	64
Gráfico N° 22 Perfiles de Clúster en Excel.....	65
Gráfico N° 23 Probabilidad por el algoritmo de Bayes.....	66

ÍNDICE DE TABLAS

Tabla N° 01. Comparación de las Metodología Kimball y Inmon.....	25
Tabla N° 02. Métodos, técnicas e instrumentos de recolección de datos.....	26
Tabla N° 03. Variables con sus indicadores.....	27
Tabla N° 04. Costos del proyecto.....	33
Tabla N° 05. Estructura de la Matriz Bus DataMart's.....	33
Tabla N° 06. Dimensiones del hecho ingreso.....	34
Tabla N° 07. Dimensiones del hecho demanda.....	34
Tabla N° 08. Requerimiento de Tecnología.....	35
Tabla N° 09. Requerimiento para la producción del DataMart.....	35
Tabla N° 10. Modelado Físico Tabla T_ dimensión concepto pago.....	41
Tabla N° 11. Modelado Físico Tabla T_ Ingresos.....	41
Tabla N° 12. Modelado Físico Tabla T_ Dimensión cliente.....	42
Tabla N° 13. Modelado Físico Tabla T_ Dimensión ubigeo.....	42
Tabla N° 14. Modelado Físico Tabla Fdemanda.....	43
Tabla N° 15. Modelado Físico Tabla T_ Dimensión área investigación.....	43
Tabla N° 16. Modelado Físico Tabla T_ Dimensión tiempo.....	44
Tabla N° 17. Modelado Físico Tabla T_ Dimensión docente.....	44
Tabla N° 18. Modelado Físico Tabla T_ Dimensión producto.....	44
Tabla N° 19. Modelado Físico Tabla T_ Dimensión carrera profesional.....	45
Tabla N° 20. Modelado Físico Tabla T_ Tiempo mes.....	45

Tabla N° 21. Flujo de datos de la tabla concepto pago hacia t_dimensión concepto pago.....	47
Tabla N° 22. Flujo de datos de la tabla ubigeo pago hacia t_dimension ubigeo.....	47
Tabla N° 23. Flujo de datos de la tabla alumno pago hacia t_dimensión cliente.....	47
Tabla N° 24. Flujo de datos de la tabla docente pago hacia t_dimensión docente.....	47
Tabla N° 25. Flujo de datos de la tabla concepto catalogo diplomado hacia t_dimensión producto.....	48
Tabla N° 26. Flujo de datos de la tabla áreas investigación hacia t_dimensio área investigación.....	48
Tabla N° 27. Flujo de datos de la tabla inscripción hacia t_demanda.....	48
Tabla N° 28. Flujo de datos de la tabla registro pago hacia t_ingresos.....	49
Tabla N° 29. hipótesis general	67
Tabla N° 30. Prueba t de la hipótesis general.....	68
Tabla N° 31. Tiempos en obtener información.....	69
Tabla N° 32. Prueba t del tiempo en obtener información.....	71
Tabla N° 33 Tiempo del ejecutivo en procesar información.....	72
Tabla N° 34. Prueba t para procesar información.....	73
Tabla N° 35. Número de errores de los reportes	74
Tabla N° 36. Prueba t para medir el número de errores de los reportes.....	75
Tabla N° 37. Escala de medición para indicadores cualitativos.....	76
Tabla N° 38. Nivel de satisfacción sobre la información solicitada.....	76
Tabla N° 39. Nivel de calidad de la información solicitada.....	77
Tabla N° 40. Resumen de Valores de los Indicadores.....	77

RESUMEN

La presente tesis se enmarca en los problemas de la falta de información por parte de los ejecutivos para saber a quién mandarles la publicidad y cuando programar los diplomados a dictarse, ya que la difusión de los diplomados se hace a través de email en forma masiva puesto que existe un desconocimiento de los perfiles de clientes, es por ello que se plantea el desarrollo de una solución de Inteligencia de Negocios utilizando la metodología de Ralph Kimball, por su base empírica y adecuada para las implementaciones de almacenes de datos; dada su gran versatilidad y enfoque. La aplicación presenta reportes analíticos como ciclo de vida de un producto, evolución de ingresos, comparativo de demanda e ingreso, así como la implementación de dashboard de demanda e ingreso para así mejorar el proceso analítico de soporte a la toma de decisiones del área comercial de la empresa educativa basado en algoritmos de bayes y clúster, y así aumentar la disponibilidad de información requerida por los ejecutivos, reducir tiempo en procesar información requerida por los ejecutivos lo que conlleva a aumentar el grado de calidad de información y con eso se aumentará el nivel de satisfacción ejecutivo de la organización. Como resultado se obtuvo una reducción de tiempo de espera en la obtención de información a un 94.75% en minutos; también mejoró la manera de cómo promocionar los diplomados con la ayuda de los algoritmos; además de la facilidad en el acceso de la información actualizada, precisa y segura alcanzando la satisfacción del ejecutivo

Palabras claves: Toma de decisiones, Inteligencia de Negocios, datamart, clúster, bayes.

ABSTRACT

The present thesis is about the problems of the lack of information that the executives have when they are planning an advanced course. The executives don't know who the potential consumers are and which the appropriate day to start the course is. In consequence the courses' information is sent by e-mail in a massive form ignoring the consumer's profile and necessities. For this reason this research proposes a Business Intelligence's solution based in a Ralph Kimbal methodology, which has an empirical and adapted base for data bases' implementations.

The Business Intelligence solutions applications presents analytical reports about the products' life cycle, the evolution of incomes, comparative of demand and revenue and the dashboard implementation of demand and revenue to improve the analytical process support in the decisionmaking commercial area, to get better results based on bayes and cluster algorithms with the purpose to increase the availability of the information required by the executives, to reduce the processing time and finally get more quality and satisfaction in the organization.

As a result, the executive have got that the waiting time of processing information decreased to 94.75% in minutes, the promotions of the courses have improved using the algorithms and also they got the access of update, accurate and secure information, being compliant and satisfied.

Keywords: decisionmaking, Business Intelligence, datamart, cluster, bayes.

I. INTRODUCCIÓN

Un estudio realizado en Europa por Information Builders Ibéric mostró el costo que tiene la falta de sistemas de toma de decisiones en las organizaciones, según estos datos, el empleado europeo medio pierde una media de 67 minutos diariamente buscando información de la compañía, lo que equivale a un 15,9% de su jornada laboral. Para una organización de 1.000 empleados que gane unos 50.000 euros al día esto equivale a 7,95 millones de euros al año de salario perdido, todo ello por la búsqueda de información para tomar una decisión.

En Chile se puede identificar niveles de aplicación del analytics en el país, de aquellos que lo ocupan, cerca del 80% de ellos pueden ser identificados como aficionados, donde a partir del uso de planillas son capaces de identificar algún nivel de reglas de operación; otro 15% pueden ser denominados semi profesionales, donde crean modelos simples, pero aún alejados desde los procesos principales; finalmente, entre un 5% están los profesionales donde están integrados a los procesos principales de las empresas, creando y utilizando modelos para la toma de decisiones.

La empresa E-knowledgegroup SAC es una empresa que se dedica al dictado de cursos de especialización entre los cuales son: los cursos presenciales y los cursos telepresenciales.

Actualmente la toma de decisiones se ve limitada por no contar con información que nos pudiera decir que es lo que está pasando con esta parte del negocio: análisis de ingresos y demanda.

Entre la diversidad de información analítica necesaria para apoyar las decisiones en el nivel operacional y administrativo podemos citar:

No saber aprovechar los hechos registrados para poder conocer las características que presentan nuestros clientes frente a la oferta de nuevos cursos. Actualmente se envía información masiva obteniendo un resultado de 30% de aceptación de los correos enviados a los distintos clientes que tiene la empresa educativa, puesto que existe un desconocimiento de los perfiles de clientes (características) para un determinado curso, otro problema que se encontró es que el sistema transaccional no contempla reportes con características analíticas, conllevando un tiempo aproximadamente de 2 días el procesar información histórica para saber por ejemplo, el ciclo de vida que tienen los cursos, la tendencia de ingresos, saber si se ha alcanzado las metas propuestas de inscripciones e ingresos, etc. También se tiene el problema de depender de una persona para la elaboración de reportes de índole analítico, para saber cómo va el negocio, si existe estacionalidad de demanda para determinados diplomados, para los ingresos.

Existe un 40% de inexactitud de información originando que la toma de decisiones se base en la experiencia (intuición) provocando una disconformidad de un 30% de los ejecutivos ya sea por los errores en que les son presentados los informes y la forma en que se les presenta lo que ocasiona un mayor tiempo para tomar alguna decisión con respecto a los diplomados que determina la empresa para promocionar a sus clientes.

Ante esta problemática resulta la siguiente pregunta ¿De qué manera ayudará la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster a mejorar los procesos analíticos del área comercial de la empresa educativa?

De esta manera se obtuvo como hipótesis que la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora los procesos analíticos en el área comercial de la empresa educativa.

Por ello la investigación tuvo como objetivo general mejorar el proceso analítico de soporte a la toma de decisiones del área comercial de la empresa educativa basado en el algoritmo de bayes y clúster. Para fortalecer el objetivo principal se han desarrollado los siguientes objetivos específicos los cuales son aumentar la disponibilidad de información requerida por los ejecutivos, reducir el tiempo de procesar información por parte del área comercial y de los ejecutivos, aumentar el grado de calidad de información para los ejecutivos, aumentar el nivel de satisfacción ejecutivo de la organización, con respecto al soporte que cuenta para la toma de decisiones.

La tesis se justifica en lo económicamente, debido a que le brindó un beneficio económico ya que mejoraron los procesos de la organización y se redujo el tiempo en la toma de decisiones.

En lo científico hubo un previo análisis donde se examinó la empresa se diagnosticó problemas, el porqué de estos y luego se decidió plantear una solución, a través del método científico logré determinar que mi solución logró cumplir los objetivos, permitiendo validar la hipótesis propuesta.

En lo tecnológico por que se plantea proponer el desarrollo de un sistema de gestión y soporte de toma de decisiones apoyados con los algoritmos de bayes y clúster aportando una propuesta novedosa, haciendo uso de la tecnología, otorgando de ésta manera la automatización de los procesos, a través de reportes y dashboards haciendo más eficiente el trabajo del de la empresa educativa

En el ámbito social, se obtuvo dos beneficiados uno de ellos fueron los internos que vienen a ser los ejecutivos debido a que se redujo el tiempo y se mejoró la aceptación de los reportes por parte de los ejecutivos y el otro beneficiado fueron los externos que vinieron a hacer los clientes quienes fueron seleccionados para poder asistir a los diplomados que fueron dictados por la institución educativa.

II. MARCO TEÓRICO

2.1 Antecedentes

Antecedente #1

Análisis, Diseño E Implementación de un datamart de clientes para el área de marketing de una entidad aseguradora.

Autor(es): Eddy Fernández Ochoa

Resumen: Este trabajo se basó en el análisis, diseño e implementación de un datamart, que contribuyó en mejorar el proceso de toma de decisiones en el área de marketing de una entidad aseguradora. Se necesitan definir indicadores para la gestión de clientes que se encuentren alineados con los objetivos del área. Asimismo, para asegurar la correcta carga se crearon nuevos procesos ETL los cuales permitirán generar información oportunamente.

Análisis de relación con la presente investigación: La elección de este antecedente se relaciona con el proyecto a desarrollar debido a que se elaboró un data mart para tener una información oportuna a la hora de la toma de decisiones

Antecedente #2

Análisis, diseño e implementación de un datamart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las direcciones de salud (2011)

Autor(es): Rosales Sedano, Carmen Pamela

Resumen:

El presente proyecto de tesis tuvo como objetivo la implementación de un datamart que permitió apoyar la toma de decisiones necesarias para cumplir con los objetivos específicos de cada estrategia sanitaria nacional dentro de las direcciones de salud. Se tomó como estrategia sanitaria piloto, la estrategia de Alimentación y Nutrición Saludable. Para ella, se realizó la captura de requerimientos, análisis, diseño y construcción del datamart. La generación de reportes por parte del datamart para cada estrategia sanitaria permite el ahorro de tiempo, pues actualmente cada estrategia debe solicitar a un área central (área de estadística) lo cual implica sobrecarga en dicha área.

Análisis de relación con la presente investigación: La elección de este antecedente se relaciona con el proyecto a desarrollar debido a que utiliza un sistema que ayuda a la toma de decisiones, implementando un datamart y se utilizó herramientas de software libre.

Antecedente #3

Análisis diseño e implementación de una solución de inteligencia de negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos. (2011)

Autor(es): Keller Gladys Rodríguez Cabañillas

Angela Lucía Mendoza Peña

El objetivo principal del proyecto fue crear escenarios de Pronósticos y reportes que apoyaron a la toma de decisiones en la compra y venta de electrodomésticos y se planteó como solución la implementación de un Datamart de Compras y Ventas. El uso de esta herramienta se tradujo en una ventaja competitiva y la

empresa se ha beneficiado por la implementación del sistema de inteligencia de negocios, utilizó la metodología Ralph Kimball o “Bottom-up” la cual resultó una solución eficaz en tiempo y recursos debido a que abarca la solución al problema en un corto plazo ayudando a reducir tiempos de respuesta en el procesamiento y análisis de información.

La presente investigación analizada guarda relación con la tesis debido a que utiliza herramientas de inteligencia de negocios para la extracción de la información en el desarrollo del datamart del área comercial el cual estará bajo la plataforma software libre de Pentaho al cual se le integran varias herramientas para cumplir con los objetivos de la tesis.

Antecedente #4

Implantación Business Intelligence en Hotel Wellington

Autor(es): Dharma Ingeniería

Resumen: En este caso de éxito el hotel Wellington tenía sus procesos de negocios resueltos ya que había implementado en modernos equipos pero faltaba en integrar su información para el proceso de toma de decisiones.

Dentro de los tres primeros meses fue posible observar los primeros resultados. El tiempo de consolidación no solo se ha reducido sino que se ha eliminado, ya que los informes se actualizan diariamente, consiguiendo que las cifras que consulta la dirección sean un reflejo real del estado diario de la empresa

Análisis de relación con la presente investigación: La elección de este antecedente se realizó por que este proyecto se hizo uso de la tecnología BI para la integración de sus áreas al igual que este proyecto de tesis pero se diferencia porque está enfocado en el área comercial de una empresa y basado en software libre para la toma de decisiones.

Antecedente # 5

El Ayuntamiento de Terrassa tomará mejores decisiones gracias a Microsoft Business Intelligence y Bismart

Autor(es): Bismart

Resumen: El objetivo del ayuntamiento ha sido afrontar los problemas que surgían tanto en la toma de decisiones como en el análisis de los datos, y esta plataforma de cuadros de mando para la administración local construida sobre Microsoft Business Intelligence les ha aportado un importante incremento en el rendimiento, así como la oportunidad de mejorar la gestión y optimizar los costes.

Análisis de relación con la presente investigación: La elección de este antecedente se realizó por que este proyecto se hizo uso de la tecnología BI de Microsoft, para la integración de sus áreas para mejorar la toma de decisiones en una empresa estatal al igual que este proyecto de tesis pero se diferencia porque está enfocado en el área comercial de una empresa privada y basado en software libre para la toma de decisiones gerenciales.

Antecedente #6

Arcor y la implementación de soluciones de BI de MicroStrategy

Autor(es): Community Manager SPPerú

Resumen: Como parte de un proyecto de actualización integral de las plataformas tecnológicas de la compañía en el mundo, se unificaron los criterios de los sistemas viejos y nuevos, se tomó la información que provenía de ambos mundos y se consolidó en el datawarehouse se desarrollaron las interfaces de donde se obtenía la información para confeccionar los reportes y demás actividades que exigía cada área y/o filial de la compañía. “Al ser el sistema de BI el único lugar donde se alojaban los datos de manera unificada, esto pasó a ser el input para alimentar otras herramientas, como las vinculadas con evaluación de costos, cálculos y pronósticos, entre muchas más y se tuvo como resultado cambios en versiones ágiles y sencillas, y también mayor eficiencia en la confección de reportes.

Análisis de relación con la presente investigación: La elección de este antecedente se realizó por que en este proyecto se hizo uso de la tecnología BI de MicroStrategy, y lo que se pretende implementar una solución de BI pero con software libre.

Antecedente #7

Guía de implementación de un modelo de inteligencia de Negocios para la gestión de la calidad en las universidades Peruanas

Autor(es): Luis Alberto Otake Oyama

Resumen: En este trabajo se propuso un modelo preliminar que implementa la sinergia de los modelos de BI y de TQM(total Quality management), tomando como punto clave la estrecha relación y las características similares que existen entre ellas. Este modelo sería aplicable principalmente en universidades latinoamericanas y sobretodo peruanas, dado las características particulares de éstas. La implementación del modelo permitiría la generación de ventajas competitivas en las universidades.

Análisis de relación con la presente investigación: La elección de este antecedente tiene relación con la tesis puesto que en proyecto se hizo uso de la tecnología BI, para la integración de sus áreas de una empresa educativa haciendo uso de un DSS (Sistema de Soporte de toma de Decisiones)

Antecedente #8

Incorporación de elementos de Inteligencia de Negocios en el proceso de Admisión y Matrícula de una Universidad Chilena.

Autor(es): Luis Fuentes Tapia
Ricardo Valdivia Pinto

Resumen: Este artículo describe un proceso orientado a la incorporación de elementos business intelligence - BI en la Universidad de Tarapacá (UTA). Con este fin, se implementó un datamart centrado en el área de Admisión y Matrícula de la Vicerrectoría Académica. Su desarrollo requirió de la realización de actividades tales como la obtención de los requerimientos del negocio, el análisis de las distintas fuentes de información interna y el desarrollo de un modelado dimensional basado en el esquema estrella de Kimball. Para la correcta implementación e integración de este repositorio de datos se debió realizar un proceso de extracción, transformación y carga (ETL). La creación de este DM permitió que los usuarios de la Vicerrectoría Académica pudieran

visualizar la información que requerían a través de herramientas de procesamiento analítico en línea (OLAP).

Análisis de relación con la presente investigación: La elección de este antecedente tiene relación con la tesis porque se hizo uso de la metodología de Kimball, se realizó un proceso de ETL y una creación de una datamart para el área comercial de la empresa educativa así como el uso de herramientas para la creación de dashboards.

2.2 Bases Teórico Científicas

2.2.1 Toma de Decisiones

Según Cohen (2005): La toma de decisiones es una de las actividades que se realiza con mayor frecuencia en el mundo de los negocios. Indicando que la importancia de la decisión empresarial va de la mano con el nivel del área en la cual se hace.

Las decisiones hoy en día juegan un papel importante ya que son parte esencial de la vida, tanto dentro como fuera del trabajo, los ejecutivos son los responsables de decidir entre dos o más opciones; por lo que la toma de decisiones abarca todo el proceso que supone sean adecuadas y eficaces, desde la liberación inicial hasta la puesta en práctica.

A continuación se presenta algunos de los fallos que León (2006), muestra sobre la manera tomar decisiones:

- ✓ Insuficiente estudio de la situación. Falta de información.
- ✓ Indecisión, falta de información completa.
- ✓ Precipitación, se trata de resolver los síntomas en vez de las causas.
- ✓ Decisiones extremadamente rápidas.
- ✓ Decisiones por intuición, sin elaborar criterios.
- ✓ Falta de previsión, de imaginación, de anticipación de riesgos.

Según Cohen (2005), también indica que existen cuatro fases en el proceso de toma de decisiones:

Fase de la inteligencia, en la cual se reconoce la existencia de un problema para cual se debe tomar una decisión.

Fase de diseño, en la cual se buscan alternativas de solución al problema encontrado en la fase de inteligencia.

Fase de selección, mediante la cual se elige la alternativa más óptima que se generó en la fase de diseño.

Fase de implantación consiste en poner en marcha y guiar a que la alternativa seleccionada.

➤ Tipos de sistema de Apoyo a las Decisiones

Cohen (2005), presenta los siguientes tipos de apoyo a las decisiones.

- ✓ **Sistemas de soporte para la toma de decisiones (DSS)**
Tiene como finalidad apoyar la toma de decisiones mediante la evaluación sistémica de diferentes alternativas o escenarios de decisión mediante el empleo de modelos y herramientas computacionales.

Su objetivo no es solucionar el problema sino apoya al proceso de la toma de decisión a los encargados de la organización.

✓ **Sistemas de información para ejecutivos (EIS)**

Está dirigido para el proceso de toma de decisiones para los altos ejecutivos de una organización, presentan información relevante y usan recursos visuales y de fácil interpretación, con el fin de mantenerse informados.

✓ **Sistemas para la toma de decisiones de grupo (GDSS)**

Su objetivo es lograr la participación de un grupo de personas durante la toma de decisiones en ambientes de anonimato y consenso, apoyando decisiones simultáneas.

✓ **Sistemas expertos de soporte para la toma de decisiones (EDSS)**

Son aquellos que permiten cargar bases de conocimiento integrados por una serie de reglas de sentido común para que diferentes usuarios las consulten, apoyen la toma de decisiones, la capacitación, etc.

➤ **Sistemas de Soporte para la Toma de Decisiones (DSS)**

Para Cohen (2005), define a los sistemas de toma de decisiones como: Un conjunto de programas y herramientas que permiten obtener de manera oportuna la información que se requiere durante el proceso de la toma de decisiones que se desarrolla en un ambiente de incertidumbre.

En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida o un ERP sofisticado. Sin embargo, no es así: estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, etc.

Para Cohen (2005), el DSS es una de las herramientas más emblemáticas del Business Intelligence ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión. Estas son algunas de sus características principales:

- Informes dinámicos, flexibles e interactivos, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- No requiere conocimientos técnicos. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos, haciendo drag&drop o drillthrough. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.

- Rapidez en el tiempo de respuesta, ya que la base de datos subyacente suele ser un datawarehouse corporativo o un datamart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información.
- Integración entre todos los sistemas/departamentos de la compañía. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa. Existe lo que se llama: integridad referencial absoluta.
- Cada usuario dispone de información adecuada a su perfil. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
- Disponibilidad de información histórica. En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio, etc.

El rol que cumple la toma de decisiones en la implementación del producto acreditable es tener el dominio propio en poder elegir o escoger que productos lanzar al mercado y a quienes está dirigida la publicidad sin margen de error para así ganar tiempo y dinero.

➤ **Tipos de Problemas de Decisión**

Para Cohen (2005), presenta distintos tipos de problemas que se dan al momento de realizar la toma de decisión:

- ✓ Problemas Estructurados. El conjunto de datos ya es conocido, la solución depende de una sola serie de pasos. La mayoría de problemas matemáticos son estructurados siempre se utiliza un algoritmo, una fórmula para poder solucionar el problema. Un ejemplo de estos problemas es la decisión de cuanto se va a producir en un determinado mes. Estas decisiones son muy fáciles de predecir, solo cambian algunas variables.
- ✓ Problemas No Estructurados. En estos tipos de problemas no hay un algoritmo que nos permita llegar a una solución en sí. Este tipo de decisiones son difíciles de tomar, ya que no se puede predecir el tipo y el escenario de la decisión o porque no hay suficiente información sobre los factores que afectan la solución.
- ✓ Problemas Semi estructurados. Estos problemas no son completamente estructurados. Estos problemas suceden cuando hay un algoritmo que puede dar un aproximado de certeza de solución sin embargo no te garantiza el cien por ciento de la solución.
- ✓ Problemas Programables. Son los problemas estructurados en los cuales se puede elaborar una solución mediante un programa. En este sentido los problemas estructurados no son programables por qué no se puede utilizar

un programa para resolverlo. Sin embargo se pueden utilizar para abordar y tratar de dar una solución a un problema no estructurado.

La presente tesis desarrollada encaja en el tipo de problema semi estructurado ya que si bien hace uso de algoritmos la decisión la toma los ejecutivos no el sistema.

2.2.2 Business Intelligence

Al respecto Microstrategy(2006), Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc.

Los principales productos de Business Intelligence que existen hoy en día son:

- Cuadros de Mando Integrales (CMI)
- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)

Según Evelson Boris, La capacidad para tomar decisiones de negocio precisas y de forma rápida se ha convertido en una de las claves para que una empresa llegue al éxito. Sin embargo, los sistemas de información tradicionales (como la mayoría de los programas de gestión, las aplicaciones a medida, e incluso los ERP más sofisticados), suelen presentar una estructura muy inflexible para este fin. Aunque su diseño se adapta con mayor o menor medida para manejar los datos de la empresa, no permite obtener la información de los mismos, y mucho menos extrapolar el conocimiento almacenado en el día a día de las bases de datos. Las principales características que limitan estos sistemas son:

- Gran rigidez a la hora de extraer datos, de manera que el usuario tiene que ceñirse a los informes predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- Necesidad de conocimientos técnicos. Para la generación de nuevos informes o métricas suele resultar ineludible acudir al departamento técnico, solicitando una consulta adecuada para interrogar la base de datos.

- Largos tiempos de respuesta, ya que las consultas complejas de datos suelen implicar la unión de tablas operacionales de gran tamaño, lo que se traduce en una incómoda espera que dificulta la fluidez del trabajo.
- Deterioro en el rendimiento del SI. Cuando la base de datos consultada, para generar informes o ratios de negocio, es la misma que la que soporta el operativo de la empresa, el funcionamiento del sistema puede degradarse hasta afectar y paralizar a todos los usuarios conectados.
- Falta de integración que implica islas de datos. Muchas organizaciones disponen de múltiples sistemas de información, incorporados en momentos distintos, para resolver problemáticas diferentes. Sus bases de datos no suelen estar integradas, lo que implica la existencia de islas de información.
- Problemas para adecuar la información al cargo del usuario. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
- Ausencia de información histórica. Los datos almacenados en los sistemas operacionales están diseñados para llevar la empresa al día, pero no permiten contrastar la situación actual con una situación retrospectiva de años atrás.

➤ **Ciclo del Business Intelligence**

Como explica Vitt E, Luchkevich M, Misner S (2003), el BI debe ser tomado como un ciclo de mejora continua que comprende:

- **Análisis.** Orientado más en el conocimiento básico de nuestro negocio: que es importante para nuestros clientes, proveedores, trabajadores, etc. Ponernos un poco en supuestos como que pasaría si decidimos subir un poco el precio a nuestros productos.
En esta etapa se analizan las mejoras para la empresa se hizo uso de entrevistas para poder ofrecer una alternativa de solución a la problemática que tenía la empresa.
- **Ideas.** Después de realizar el análisis y con las herramientas del BI pueden surgir ideas de mejoras para el negocio.
En esta etapa los analistas del negocio toman nuevas propuestas para poder mejorar el producto que se está lanzando al mercado para una mayor aceptación del cliente.
- **Acción.** Está relacionado con el proceso de toma de decisiones. Esta acción respaldada por el análisis hecho anteriormente va a permitir que la acción tenga una mayor aceptación por parte de la empresa.
Esta etapa tiene que ver con el estudio realizado y haber mandado la publicidad a los posibles clientes que se inscribirán en el diplomado que está realizando la empresa.
- **Medición.** El BI permite fijar los estándares, niveles o indicadores claves de gestión (KPI) para hacer un seguimiento de los resultados realimentando a

cada área del negocio. Un ejemplo de indicadores de gestión pueden ser: ventas hechas por un empleado, precio de venta promedio.

En esta parte se hace una medición de manera estadística para ver los productos que más movimientos tuvieron en el transcurso del año ya sean en ingresos y demandas.

- **Datamart**

Al respecto INEI (1997), Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio datawarehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento.

Se vio conveniente la creación de un datamart ya que es una empresa pequeña y es una solución para problemas inmediatos y sus consultas son más rápidas, su costo es más económico porque las herramientas son de software libre

- **Bases de datos OLTP y OLAP**

OLTP - On-Line Transactional Processing

Según Vitt, Elizabeth 2000. Señala que un sistema OLTP está diseñado para administrar datos originales o datos en crudos, los cuales requieren eficiencia y procesamiento de transacciones al minuto y al más bajo nivel de detalle.

Los OLTP son base de datos orientados al procesamiento de transacciones, generando un proceso atómico una transacción las mismas que pueden involucrar operaciones de inserción, modificación y borrado de datos.

Por otro lado, señalar que el proceso transaccional es típico de las bases de datos operacionales.

La base empresa educativa ya cuenta con un base de datos OLTP de la cual se han extraído los datos para poder crear el datamat para el apoyo en la toma de las decisiones.

OLAP - On-Line Analytical Processing

Según Vitt, Elizabeth 2000. Señala que un OLAP permite análisis multidimensional compuesto por herramientas de interface y estructura de bases de datos que permitan acceso instantáneo y una fácil manipulación por parte del usuario; asimismo proporciona un modelo de datos intuitivos y conceptual para los usuarios que no tengan experiencia como analistas puedan comprender y rápidamente relacional.

La base de datos OLAP, la implementación del cubo del cual se van a ir procesando los datos para obtener la información necesaria para la toma de decisiones de la empresa educativa.

2.2.3 Algoritmo de Bayes

Para Microsoft (2012) Naive Bayes es una técnica de clasificación y predicción que construye modelos que predicen la probabilidad de posibles resultados. Naive Bayes utiliza datos históricos para encontrar asociaciones y relaciones y hacer predicciones.

Datos requeridos para los modelos Bayes naive

Al preparar los datos para su uso en un modelo de entrenamiento Bayes naive, conviene comprender qué requisitos son imprescindibles para el algoritmo, incluidos el volumen de datos necesario y la forma en que estos datos se utilizan.

Los requisitos para un modelo Bayes naive son los siguientes:

Una columna de una sola clave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Columnas de entrada: en un modelo Bayes naive, todas las columnas deben ser discretas o discretizadas.

En un modelo Bayes naive, también es importante asegurarse de que los atributos de entrada sean independientes unos de otros. Esto es particularmente importante al utilizar el modelo para la predicción.

El motivo es que, si utiliza dos columnas de datos que ya están estrechamente relacionadas, el efecto sería multiplicar la influencia de esas columnas, lo que puede ocultar otros factores que influyen en el resultado.

Al contrario, la capacidad del algoritmo de identificar las correlaciones entre las variables es útil cuando está explorando un modelo o conjunto de datos, para identificar las relaciones entre las entradas.

Al menos una columna de predicción: el atributo de predicción debe contener valores discretos o discretizados.

Los valores de la columna predecible se pueden tratar como entradas. Este ejercicio puede ser útil si explora un nuevo conjunto de datos, para encontrar relaciones entre las columnas.

Atributos	Estados	Población... Tamaño: 18484	0 Tamaño: 9352	1 Tamaño: 9132	ausente Tamaño: 0
Age	<ul style="list-style-type: none"> ● 38 - 43 ● 29 - 34 ● 43 - 48 ● Other 				
Commute Distance	<ul style="list-style-type: none"> ● 0-1 Miles ● 2-5 Miles ● 1-2 Miles ● Other 				
Education	<ul style="list-style-type: none"> ● Bachelors ● Partial College ● High School ● Other 				
Marital Status	<ul style="list-style-type: none"> ● M ● S ● Missing 				
Number Cars Owned	<ul style="list-style-type: none"> ● 2 ● 1 ● 0 ● Other 				
Number Children At Home	<ul style="list-style-type: none"> ● 0 ● 1 ● 2 ● Other 				
Occupation	<ul style="list-style-type: none"> ● Professional ● Skilled Manual ● Management 				

Gráfico 01: Visor de Bayes Naives
Fuente: Microsoft

2.2.4 Algoritmo de Clúster

Para Vicente (2010) El algoritmo utiliza técnicas iterativas para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

Los modelos de agrupación en clústeres identifican las relaciones en un conjunto de datos que no se podrían derivar lógicamente a través de la observación casual.

También es conocido con el nombre clasificación no supervisado.

Cuando el número de columnas es bastante grande se pueden formar también grupos de columnas con similar comportamiento y en consecuencia se puede reducir la dimensionalidad que es muy conveniente si se quiere usar luego un modelo para hacer predicciones.

Podemos encontrarnos dos tipos fundamentales de métodos de clasificación:

Jerárquicos y No Jerárquicos. En los primeros, la clasificación resultante tiene un número creciente de clases anidadas mientras que en el segundo las clases no son anidadas.

Los métodos pueden dividirse en aglomerativos y divisivos. En los primeros se parte de tantas clases como objetos tengamos que clasificar y en pasos sucesivos vamos obteniendo clases de objetos similares, mientras que en los segundos se parte de una única clase formada por todos los objetos que se va dividiendo en clases sucesivamente.

Por otro lado Microsoft (2012) nos dice. El algoritmo de clústeres identifica primero las relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas. Un gráfico de dispersión es una forma útil de representar visualmente el modo en que el algoritmo agrupa los datos.

El gráfico de dispersión representa todos los casos del conjunto de datos; cada caso es un punto del gráfico. Los clústeres agrupan los puntos del gráfico e ilustran las relaciones que identifica el algoritmo.

Después de definir los clústeres, el algoritmo calcula el grado de perfección con que los clústeres representan las agrupaciones de puntos y, a continuación, intenta volver a definir las agrupaciones para crear clústeres que representen mejor los datos. El algoritmo establece una iteración en este proceso hasta que ya no es posible mejorar los resultados mediante la redefinición de los clústeres.

2.2.5 Metodologías de sistemas de Inteligencia de Negocios.

- **Metodología de Bill Inmon.**

Según (Inmon 1992) propone mecanismos para llevar a cabo la correcta realización de un Data Warehouse, en donde el diseño comienza con la introducción de los datos debido a las grandes cargas de datos que deben hacerse antes de su introducción al Data Warehouse, Inmon asocia frecuentemente los Data Warehouse a nivel empresarial, que involucran desde un inicio todo el ámbito corporativo, sin centrarse en un incremento específico hasta después de haber terminado completamente el diseño del Data Warehouse. En su filosofía, un Data Mart es sólo una de las capas del Data Warehouse y los Data Mart son dependientes del depósito central de datos o Data Warehouse Corporativo y por lo tanto se construyen después de él.

- **Metodología de Ralph Kimball**

Microstrategy (2006) El marco presentado por Ralph Kimball con el nombre de Business Dimensional Lifecycle (BDL) ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Los pasos a seguir son los siguientes:

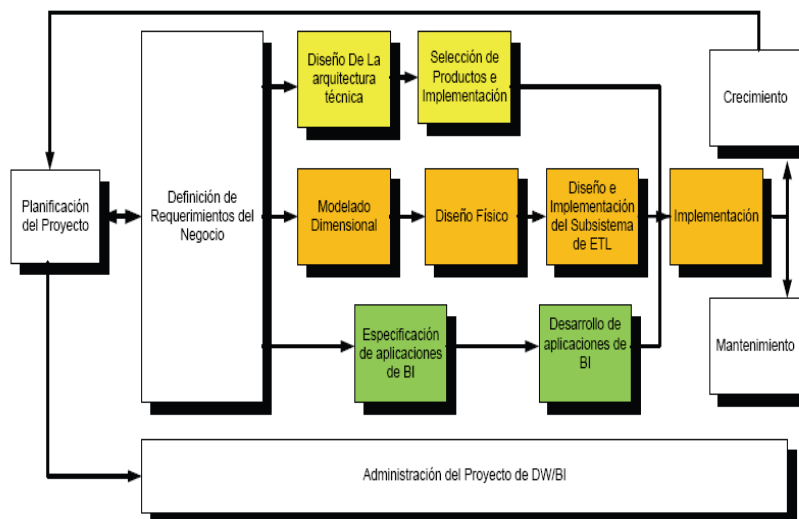


Gráfico N° 02. Business Dimensional Lifecycle (BDL) Ralph Kimball
Fuente: Microstrategy (2006)

Se pueden apreciar los distintos criterios que hay entre las metodologías de Kimball y la de Bill Inmon's.

Criterio	KIMBALL	INMON
Paradigma	DWH es el conglomerado de todos los Data Marts dentro de la empresa.	DWH es una parte del sistema global de BI de una empresa. Los Data Marts obtienen su información del DWH.
Almacenamiento	La información está siempre almacenada en el modelo dimensional.	La información se almacena en 3era forma normal.
Enfoque	Bottom-Up	Top-Down
Tiempo (implementación)	Tiempos cortos por cada Data Mart, se puede extender si no se planea adecuadamente la posterior integración de los mismos.	Requiere mucho más tiempo ya que se implementa el DWH en su totalidad.
Criterio	Kimball	Inmon
Costos	Costos reducidos por implementación independiente de los Data Marts.	Costos elevados por la cantidad de data a implementar.
Modelo	Identificación de dimensiones y hechos.	Tres niveles en el modelo: Alto Nivel ERD, (Entity Relationship) Nivel Medio DIS, (Data Item Set) Nivel Bajo (Modelo Físico).

Estructuras	Estrella	Copo de nieve, tablas puente.
--------------------	-----------------	--------------------------------------

Tabla 01: Comparación de las Metodología Kimball y Inmon

2.2.6 Aporte de las aplicaciones de Inteligencia de Negocios a la Tesis

El aporte relevante que se presenta en la tesis desarrollada es de que gracias a la inteligencia de negocios y los algoritmos utilizados la empresa podrá disponer de información más relevante, fidedigna y actualizada, estará en mejor posición para tomar decisiones que se ajusten a la realidad de la empresa y del mercado, la empresa también podrá empezar a plantearse proyectos que antes no lo podían hacer por falta de información fiable. En general, disponer de información relevante ampliará su horizonte, ya que el conocimiento es poder. Por otro lado se mejorará la capacidad de respuesta, se dispondrá más rápidamente de la información necesaria de una manera más visual e intuitiva que ayudará en la toma de decisiones por parte de los ejecutivos, y para terminar se mejoró la visibilidad de los formatos para su mayor comprensión permitiéndoles mejorar la perspectiva y la habilidad para comprender el negocio.

III. MATERIALES Y MÉTODOS

3.1 Diseño de investigación

TIPO DE INVESTIGACIÓN:

- Tecnológica Aplicada, puesto que se va a mejorar el soporte para la toma de decisiones del área comercial de la empresa educativa basado en algoritmos de bayes y clúster,

FORMULACIÓN DE LA HIPÓTESIS

La implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora los procesos analíticos del área comercial de la empresa educativa.

POBLACIÓN, MUESTRA DE ESTUDIO y MUESTREO

La población para el presente trabajo de investigación se constituyó por las personas que toman decisiones en la empresa E-knowledgegroup S.A.C.:

La población se determinó de la siguiente manera:

Por ser la población pequeña en este caso: 6 personas que forman la junta directiva el error es del 0% y la muestra serían las personas que conforman la directiva es decir (3 personas).

MÉTODOS, TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Técnica	Justificación	Instrumento	Aplicado en
Observación	Permitirá tener una perspectiva de la situación problemática, mediante la cual nos pudimos dar cuenta de la realidad de la empresa y como se realiza el proceso de toma de decisiones.	Observación	E- Knowlegdegrou SAC
Entrevista	Para la realización de la entrevista, se prepararon preguntas enfocadas al proceso de toma de decisiones en la empresa. Realizando para ello una investigación sobre la información necesaria que ayudan a las decisiones del tipo táctico en la empresas.	Hojas de apuntes	E- Knowlegdegrou SAC

Tabla 02: Métodos, técnicas e instrumentos de recolección de datos

PLAN DE PROCESAMIENTO PARA ANÁLISIS DE DATOS

Una vez realizada la recolección de datos a través de entrevistas y observación, hay que validar los datos.

Validación y Edición: En esta etapa se analiza si se hicieron las preguntas adecuadas y suficientes para la obtención de los datos. Si se cometió errores por parte del entrevistador o por parte del entrevistado

Variable	Dimensión	Indicador	Instrumento	Operacionalización
Proceso analítico de soporte a la toma de decisiones	Disponibilidad Información	Conocimiento de los ejecutivos con respecto a la información disponible	Encuesta	$\frac{\text{Nº de Ejecutivos B/MB}}{\text{Nº de Ejecutivos Encuestados}}$
	Tiempo para procesar información del área comercial	Tiempo en obtener información solicitada por ejecutivos	Cronómetro	Nº de minutos que se toma el trabajador en realizar las transacciones encomendadas por el ejecutivo
	Tiempo procesar información Ejecutivo	Tiempo del ejecutivo en procesar información	Cronómetro	Nº de minutos del ejecutivo en procesar la información solicitada
	Exactitud de información	Porcentaje de inexactitud de información	Conteo	$\frac{\text{Nº de errores con sistema tradicional} - \text{Nº de errores con sistema implementado}}{\text{Nº de Ejecutivos B/MB}}$
	información Formato	Porcentaje de aceptación de formato de información	Encuesta	$\frac{\text{Nº de Ejecutivos B/MB}}{\text{Nº de Ejecutivos Encuestados}}$
	Nivel de satisfacción Ejecutivo	Porcentaje de ejecutivos	Encuesta	$\frac{\text{Nº de Ejecutivos B/MB}}{\text{Nº de Ejecutivos Encuestados}}$

Tabla 03: Variables con sus indicadores

3.2 Metodología

✓ Planificación

Esta etapa se encarga de identificar y definir el alcance del proyecto así como la justificación del mismo, realizando un análisis de factibilidad.

En la parte de planeación se tiene en cuenta al personal y los equipos con los que se trabajará.

Se focaliza sobre recursos, perfiles, tareas, duraciones y secuencialidad.(cronograma)

✓ Análisis de requerimientos

Un factor determinante en el éxito de un proceso de DW es la interpretación correcta de los diferentes niveles de requerimientos, expresados por los diferentes niveles de usuarios.

Los diseñadores de los DW deben entender los factores claves que guían al negocio para determinar efectivamente los requerimientos.

El éxito del proyecto depende de una comprensión sólida de las necesidades de negocio.

✓ Modelado dimensional

En esta etapa se definen las dimensiones y atributos con las que se va a trabajar, también la arquitectura de construcción, los esquemas modelos y las tablas de hechos describiendo así el concepto de DataMart's

✓ Diseño físico

El diseño físico de las base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico.

Preparan el entorno de base de datos.

✓ Diseño de presentación de datos

Las principales sub etapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga (ETL process).Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico acordado. Los procesos de transformación se define como los procesos para convertir o recodificar los datos fuente a fin poder efectuar la carga efectiva del Modelo Físico. Los procesos de carga de datos son los procesos requeridos para poblar el Data Warehouse.

✓ Diseño de la arquitectura técnica

Se debe tener en cuenta tres factores: Los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de DW. Es el marco arquitectural completo del proyecto.

✓ **Selección del producto e instalación**

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura cómo será la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc.

✓ **Especificación de aplicaciones para los usuarios finales**

Esta etapa se delimita la aplicación desarrollada ya que todas las personas que accedan a esta, no tendrán los mismo roles o perfiles necesarios para acceder a la información del negocio necesario para la toma de decisiones. Destacando cuatro pasos principales:

Determinación del conjunto de templates iniciales (identificar reportes candidatos, clasificarlos y priorizarlos)

Diseño de la estrategia de navegación dentro de la aplicación (esquema de pantallas, esquema de carpetas – directorios, criterios de agrupamiento -por datos, por dueño, por regla del negocio, etc.)

Determinación de estándares (nombre de objetos, ubicación de objetos, formato de las salidas)

Detalle de las especificaciones (definición: nombre, descripción o propósito, frecuencia, parámetros, restricciones, layout, etc.)

✓ **Desarrollo de aplicaciones para usuarios finales**

Esta etapa es la continuación de la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de la metadata y construcción de reportes específicos.

✓ **Implementación**

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio.

✓ **Mantenimiento**

Cuando el sistema está en producción Incluye:

Tareas técnico operacionales que son necesarias para mantener el sistema operando óptimamente.

Monitorio del uso.

Tuning del desempeño.

IV. RESULTADOS

4.1 Planeamiento del proyecto

4.1.1 Identificación del escenario del Proyecto

Reseña Histórica

Eknowledge Group S.A.C inicia sus actividades académicas el día 02 de Junio del 2006 con su fundador el ingeniero Edward Haro Maldonado quien a través de su sello particular, transformó a la institución en un aporte real al estudio. Nuestro grupo ha ido creciendo en forma progresiva motivando que su infraestructura y personal docente también se incremente, al igual que otros servicios.

En la actualidad la institución Eknowledge Group S.A.C cuenta con dos formas de aprendizaje ya sea de forma presencial y telepresencial, los diplomados que se dictan son para todas las carreras profesionales que permite considerarla como una institución educativa seria en el norte del país, acorde con las exigencias educativas actuales.

➤ **Razón Social**

Eknowledge Group S.A.C

➤ **RUC**

20539153961

➤ **Misión**

Somos una red de capacitación virtual con nodos en diferentes lugares del Perú, que brinda conocimiento especializado a profesionales de distintas áreas del conocimiento, así como a emprendedores y empresarios, de manera práctica y responsable, contamos para ello con una estructura acorde y con docentes altamente especializados, mediante plataformas tecnológicas síncronas y asíncronas, contamos con personal de calidad, identificado con el cliente y altamente calificado tanto en el ámbito tecnológico como pedagógico.

➤ **Visión**

Seremos la red de capacitación virtual en conocimientos orientados a profesionales y emprendedores más grande y prestigiosa del Perú, con una imagen ganada y con un alto valor de marca. Nos vemos además integrando ciudades al interior del Perú fomentando así la inclusión de la educación y el conocimiento especializado, compartiendo experiencias aplicativas con países y personas hispanohablantes.

4.1.2 Identificación de los Usuarios y áreas intervinientes

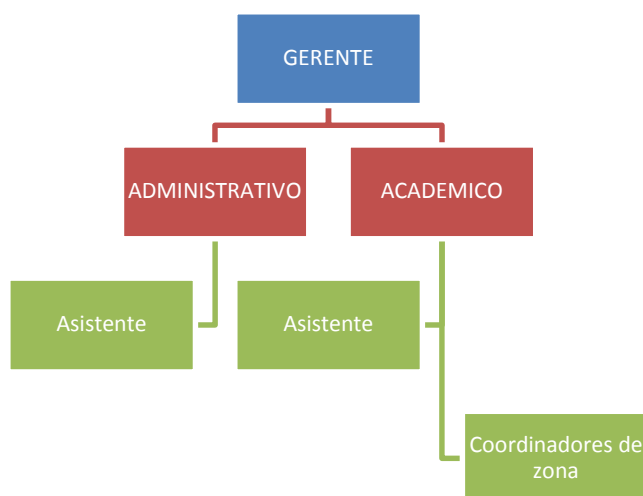


Gráfico N° 03. Organigrama de la empresa Eknowledge Group S.A.C

Fuente: Eknowledge Group S.A.C

Este organigrama nos muestra cómo está dividido la empresa educativa Eknowledge Group S.A.C

GERENTE

Esta área se encuentra a cargo Ing. Edward Haro Maldonado y se encarga de la dirección, de la empresa y de la toma de decisiones a nivel de la organización.

ADMINISTRATIVO

Encargado de las partes legales y de los ingresos de la empresa.

ACADEMICO

Es el encargado de planificar los diplomados que se van a emitir durante el año., capacitaciones a las instituciones públicas y privadas, intervienen en el apoyo de las tomas de decisiones al gerente de la institución.

ASISTENTES

Son los encargados de apoyar en las actividades académicas, administrativas según el área donde este ubicados.

4.1.3 Análisis de Factibilidad

El estudios de factibilidad es la base que permite servir de apoyo a la persona que invertirá para tomar una decisión en lo referente a que si se debe o no continuar con el proceso de desarrollo del proyecto de BI, este estudio profundiza la investigación en fuentes secundarias y primarias detallando la tecnología que se empleara determinando los costos totales y la rentabilidad económica del proyecto.

Factibilidad Operativa

Para realizar el análisis de factibilidad Operacional del empresa será necesario capacitar a las personas encargadas del manejo del DSS, en este caso la gerencia y el jefe de área académico, para que puedan hacer un mejor análisis y por ende tomar mejores decisiones.

Cada persona responsable de las áreas con las que cuenta la empresa se verá involucrado con el DSS ya que ellos podrán también conocer la evolución de los diferentes informes o reportes que arroje el sistema.

El DSS contara con una interfaz Web que permitirá que el usuario pueda acceder fácilmente a toda la información necesaria para la toma de decisiones en sus respectivas áreas.

La actualización de la base de datos dimensional se realizará automáticamente.

Factibilidad Técnica

Para realizar el análisis de factibilidad técnica de la empresa se tiene que cuenta actualmente con equipos informáticos, equipos de red, comunicación y también cuenta el servidor de base de datos y web necesario para poder desarrollar la aplicación planteada.

Factibilidad Económica

Un aspecto interesante de la implementación de un Datamart es la justificación del costo sobre la base del retorno de la inversión. Tal análisis requiere que los beneficios del Datamart sean conocidos y cuantificados antes de comenzar con su construcción.

Cuando se menciona la reducción de costos se refiere a:

- Costos de tomar malas decisiones por no contar con la información adecuada en el momento oportuno.
- Costos de programación de nuevas aplicaciones o adaptaciones de las existentes.

Análisis de Costos

Se aplicaran los costos necesarios para la realización del proyecto, tanto los gastos que se realizan en material de oficina como los costos de implementación y operación.

Rubro	Cantidad	Unidad de medida	Precio Unitario. (S/.)	Total. (S/.)
Bienes materiales				
Papel Bond	1	Millar	s/.24.00	s/.24.00
Lapiceros	3	Unidades	s/./3.00	s/./9.00
Tinta de impresora negro	1	Litro	s/./40.00	s/./40.00
Tinta de impresora color	3	Litro	s/./50.00	s/./150.00
Memoria USB	1	8 Gb	s/./30.00	s/./30.00
Folder Manila	20	Paquete	s/./0.70	s/./14.00
Cds	3	Unidades	s/./1.50	s/./4.50
				s/./271.50
Servicios				
Internet	6	Meses	s/./85.00	s/./510.00
Energía eléctrica	6	Kwh/mes	s/./120.00	s/./720.00
Espiralado	3	Unidad	s/./3.00	s/./9.00
Movilidad	2	Global	s/./25.00	s/./50.00
Licenciamiento de SQL SERVER 2008 Estándar Edition	1	Unidad	s/./18989.47	s/./18989.47
				s/./20821.47
Total del Presupuesto del proyecto				s/./20821.47

Tabla N° 04. Costos del proyecto

4.2 DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO

4.2.1 Requerimientos del Negocio

4.2.1.1 Requerimientos de Usuarios Finales

Para realizar la recolección de requerimientos se realizaron entrevistas a los encargados de las diferentes áreas de la empresa, a través del levantamiento de información de los documentos utilizados nos permitió identificar lo siguiente.

HECHO	Medida	Concepto Pago	Ubigeo	Cliente	Tiempo	Producto	carrera Profesional	Docente	área Investigación
INGRESO	Monto Total	X	X	X	X	X			
DEMANDA	Cantidad inscritos		X	X	X	X		X	X

Tabla N° 05. Estructura de la Matriz Bus DataMart's

Requerimientos de análisis de ingresos.

Este hecho es necesario analizar bajo las siguientes perspectivas (dimensiones).

Perspectivas	Descripción
Concepto pago	Nos permite tener una idea de la distribución de ingreso por los diferentes conceptos (matriculas, pensiones, costo de certificado).
Ubigeo	Esta dimensión nos permite saber el ingreso que tiene la institución por las diferentes zonas geográficas del Perú.
Cliente	La dimensión cliente nos permite analizar el ingreso que generan los clientes con los que cuenta la empresa.
Tiempo	La dimensión tiempo nos permite saber los ingresos que hay en los diferentes periodos del tiempo (mes, trimestre, semestre), así como la tendencia de ingresos, así comparar los ingresos en diferentes periodos de tiempo.
Producto	esta dimensión podremos conocer los ingresos que genera cada uno de los productos, por área y la carrera profesional a la cual pertenece

Tabla N° 06. Dimensiones del hecho ingreso.

Requerimientos de análisis de demanda.

El hecho demanda es necesario analizar bajo las siguientes perspectivas (dimensiones).

Perspectivas	Descripción
Ubigeo	Esta dimensión me permite saber la demanda que se tiene por las diferentes zonas geográficas del Perú.
Cliente	Esta dimensión me permite analizar la cantidad de veces que un cliente se inscribe en los diferentes diplomados emitidos por la empresa.
Tiempo:	Esta dimensión me permite saber la cantidad de inscritos que hay en los diferentes periodos del tiempo (mes, trimestre, semestre), así como la tendencia de inscripciones, comparar los inscritos en diferentes periodos de tiempo.
Producto:	A través de esta dimensión podremos conocer la demanda que genera cada uno de los productos, por área y la carrera profesional a la cual pertenece.

Tabla N° 07 Dimensiones del hecho demanda.

El haber considerado todas estas perspectivas de análisis para los hechos nos permitirá también analizar los hechos con la participación de una, algunas o todas estas dimensiones.

Contamos con las siguientes fuentes de información:

- ✓ Analizar Reportes de Gestión
- ✓ Revisar Base de Datos

Analizar Reportes de gestión

La revisión de los reportes de gestión nos permitió encontrar:

- ✓ Medidas (manera como se cuantifica un hecho)
- ✓ Dimensiones

Requerimientos del Desarrollador

- ✓ **Requerimientos de Tecnología**

REQUERIMIENTOS DE TECNOLOGÍA	ELEMENTOS
Servidor para DataMart's	PC Intel Core i5- 3.2 GHZ – RAM 8 GB, HD Sata 500 GB y adaptador de red
PC - Estación de trabajo	PC Pentium IV Intel – 2.8 GHZ – RAM 512 MB HD 120 GB y adaptador de red
Red LAN	Switch D-LINK 10/100 de 24 puertos y cable UTP Cat 5e
Sistema Operativo	Microsoft Windows 7
Gestor de Base de Datos	Postgresql
Software de Desarrollo	TALEND OPEN STUDIO SCHEMA WORK BEANCH MONDRIAN
Protocolo de Comunicación	TCP/IP

Tabla N° 08. Requerimiento de Tecnología

- ✓ **Requerimiento de disposición para la producción de DataMart**

REQUERIMIENTOS DE DISPOSICIÓN	ELEMENTOS
Precisión de la Información	100% precisa
Seguridad	Solo alta gerencia y usuarios autorizados
Desempeño del Sistema	En tiempo real
Mantenimiento y Actualización de datos	Diario, en el momento en que se finaliza la jornada laboral.
Administración del Almacenamiento	Verificar datos consistentes y realizar constantes depuraciones de LOG para lograr más espacio en disco
Comunicación y Conectividad	Disponible permanentemente las 24 horas del día

Tabla N° 09. Requerimiento para la producción del DataMart

4.3 Modelo Dimensional

4.3.1 Análisis de Medidas y Dimensiones

Según la información analizada se encontraron las siguientes medidas y dimensiones

4.3.2.- Star Net para el análisis.

A continuación se presenta el star net de los distintos hechos que se han utilizado para el análisis.

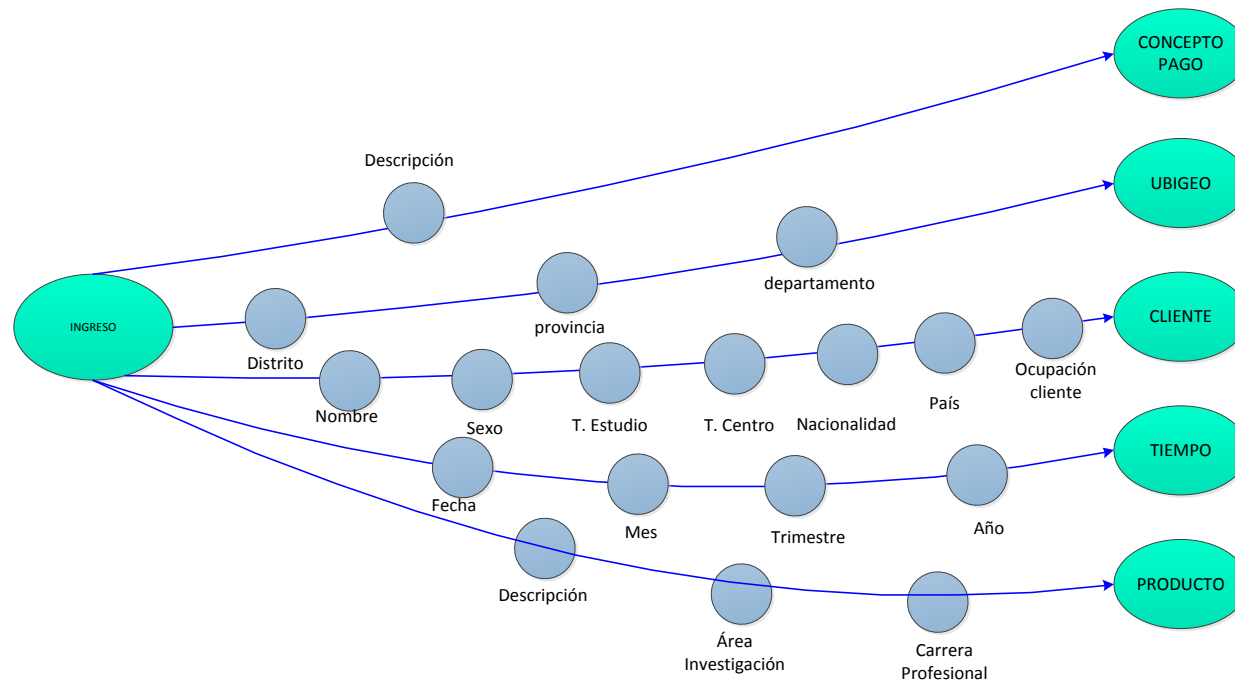


Gráfico N°04. Star net ingreso

Características:

- Análisis de ingresos por concepto pago (descripción: matrícula, pensión, certificado)
- Análisis geográfico de ingreso por distrito, provincia y departamento.
- Análisis de ingresos por tipo de cliente. El tipo se refiere la ocupación del cliente.
- Análisis de ingresos por fechas.
- Análisis de ingresos por productos ofrecidos por la empresa.

Consideraciones.

- Comparativo de ingresos entre unidades de tiempo.
- Suma acumulada de ingresos.
- Ingresos reales vs proyectados por Empresa - Producto-mes

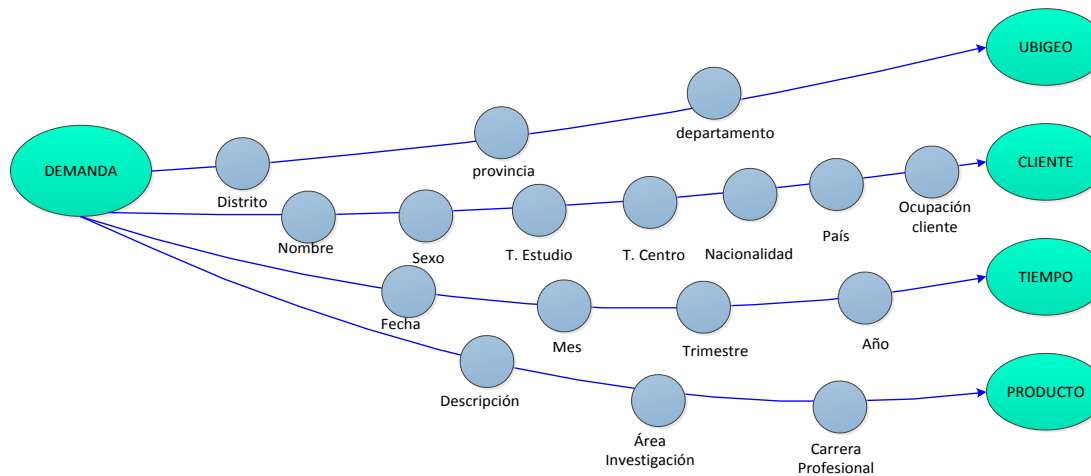


Gráfico N°05. Star net demanda

Características:

- Análisis geográfico de demanda por distrito, provincia y departamento.
- Análisis de demanda por tipo de cliente. El tipo se refiere la ocupación del cliente.
- Análisis demanda por fechas.
- Análisis de ingresos por productos ofrecidos por la empresa.

Consideraciones.

- Comparativo de demanda entre unidades de tiempo.
- demanda reales vs proyectados por Empresa - Producto-mes

4.3.3 Análisis de las Dimensiones y Jerarquías**Jerarquías**

La ventaja de considerar jerarquías para las dimensiones reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

Dimensión Tiempo

Nombre: por fecha

* año

** trimestre

*** mes

**** fecha

Dimensión Ubigeo

Ubicación geográfica

* departamento

** provincia

*** distrito

dimensión Cliente

cliente por ocupación

*ocupación

**nombre cliente

Dimensión Producto

Por carrera profesional

* carrera

** área de investigación

*** producto

Dimensión Área de investigación
Áreas por carrera profesional
*Carrera profesional
**Área de investigación

Dimensión Docente
Docente por carrera
*carrera profesional
**nombre docente

Dimensión tiempo
Por calendario
*año
**trimestre
***mes
****fecha

4.4 Modelo de datos dimensional

En esta etapa de la metodología se definirá el diseño lógico de cada una de las dimensiones a utilizarse en el desarrollo del proyecto, basándonos en los requerimientos de decisión de la empresa.

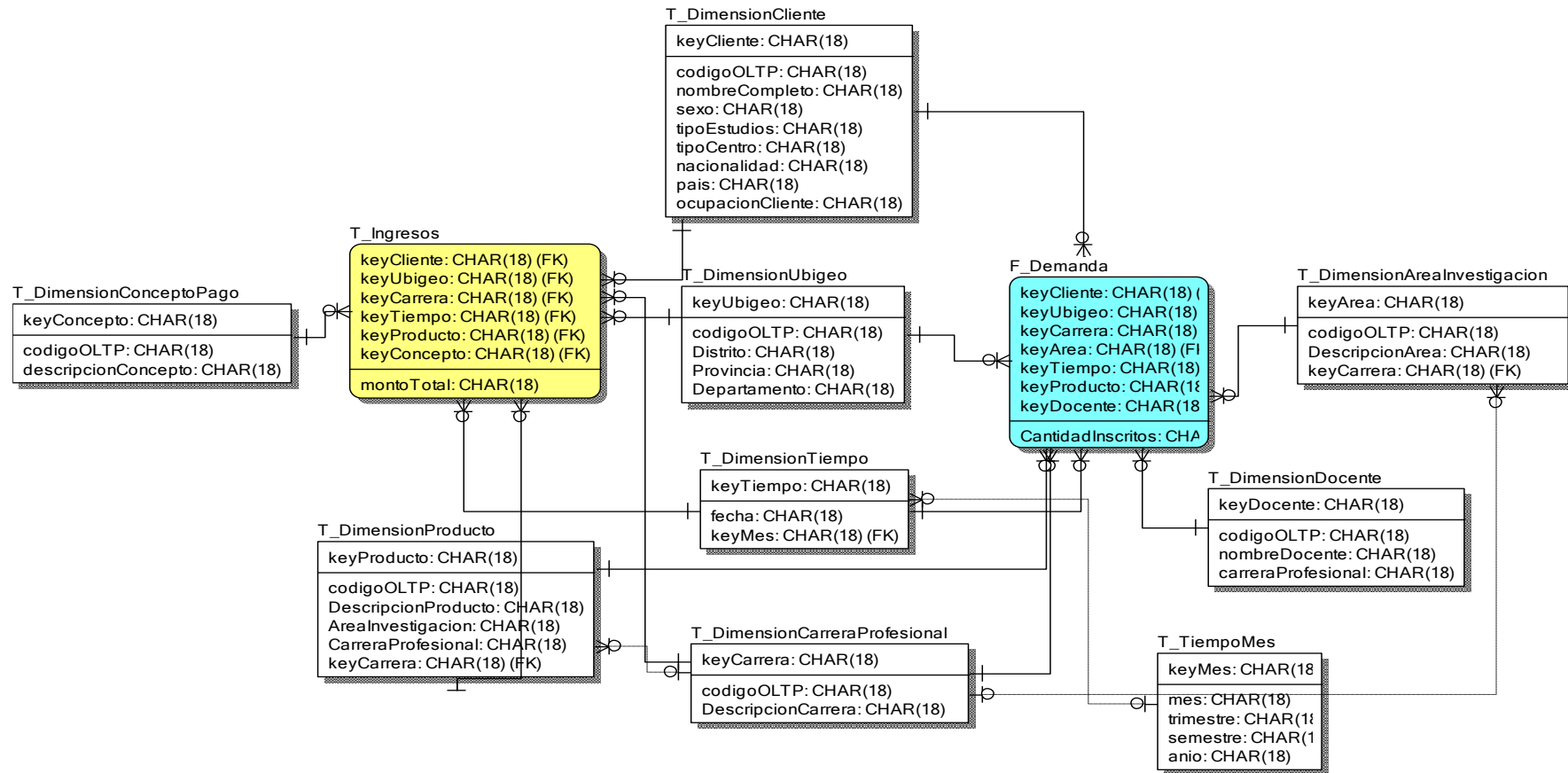


Gráfico N° 06 Diseño de Base de Datos Dimensional.

4.5 DISEÑO FÍSICO

Para el diseño físico de datos se destaca el modelo físico de la base de datos dimensional de la base de datos del Sistema Soporte a Toma de Decisiones.

T_DIMENSION CONCEPTO PAGO		DIMENSION QUE ALMACENA EL CONCEPTO DE PAGO				
Campos	Tip o de dat o	Tamañ o	Nulida d	Restricció n	Descripció n de la restricción	Descripció n del campo
keyconcepto	char	18		primary key		
codigoOLTP	char	18	no			
descripcionconcepto	char	18	not null			

Tabla N° 10. Modelado Físico Tabla T_dimension concepto pago

T_INGRESOS		TOTAL DE INGRESOS DE LOS PRODUCTOS				
Campos	Ti po de dat o	Tama ño	Nulid ad	Restricción	Descripc ion de la restricci ón	Descripc ión del campo
Keycliente	char	18	not null	Foreign key(t_dimensioncliente)		codigo del cliente
Keyubigeo	char	18	not null	Foreign key(t_dimensionubigeo)		codigo de ubigeo
Keycarrera	char	18	not null	Foreign key(t_dimensioncarrerap rofesional)		codigo de carrera profesion al
Keytiempo	char	18	not null	Foreign key(t_dimensiontiempo)		codigo de tiempo
Keyproducto	char	18	not null	Foreign key(t_dimensionproducto)		codigo de producto
keyconcepto	char	18	not null	Foreign key(t_dimensionconcepto pago)		
Montototal	char	18	not null			monto total del producto
montopagado	char	18	not null			monto pagado del producto
Montosaldo	char	18	not null			saldo del monto del producto
Cantidadocurrencias	char	18	not null			

Tabla N° 11. Modelado Físico Tabla T_Ingresos

T_DIMENSION CLIENTE DIMENSION QUE ALMACENA LOS DATOS DEL CLIENTE						
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keycliente	char	18	not null	primary key		codigo de cliente
codigoOLTP	char	18	not null			
nombrecompleto	char	18	not null			nombre del cliente
Sexo	char	18	not null			sexo del cliente
tipoEstudios	char	18	not null			
tipoCentro	char	18	not null			
nacionalidad	char	18	not null			nacionalidad del cliente
País	char	18	not null			país del cliente
ocupacioncliente	char	18	not null			ocupación del cliente

Tabla N° 12 Modelado Físico Tabla T_Dimensión cliente

DIMENSION UBIGEO DIMENSION QUE ALMACENA EL UBIGEO DE LOS CLIENTES						
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keyubigeo	char	18	not null	primary key		codigo de ubigeo
codigoOLTP	char	18	not null			
Distrito	char	18	not null			nombre del distrito donde vive el cliente
Provincia	char	18	not null			nombre de la provincia donde vive el cliente
Departamento	char	18	not null			nombre del departamento donde vive el cliente

Tabla N° 13 Modelado Físico Tabla T_Dimensión ubigeo

FDEMANDA DEMANDA DE LOS PRODUCTOS QUE SE OFRECEN						
Campos	Tipo de dato	Tamaño	Nullidad	Restricción	Descripción de la restricción	Descripción del campo
Keycliente	char	18	not null	Foreign key(t_dimensioncliente)		codigo del cliente
Keyubigeo	char	18	not null	Foreign key(t_dimensionubigeo)		codigo de ubigeo
Keycarrera	char	18	not null	Foreign key(t_dimensioncarreraprofesional)		codigo de carrera profesional
Keytiempo	char	18	not null	Foreign key(t_dimensiontiempo)		
keyproducto	char	18	not null	Foreign key(t_dimensionproducto)		codigo del producto
keydocente	char	18	not null	Foreign key(t_dimensiondocente)		codigo del docente
Cantidadinscritos	char	18	not null			cantidad de clientes inscritos en el diplomado

Tabla N° 14. Modelado Físico Tabla Fdemanda

T_DIMENSION AREA INVESTIGACION DIMENSION QUE ALMACENA EL AREA DE INVESTIGACION						
Campos	Tipo de dato	Tamaño	Nullidad	Restricción	Descripción de la restricción	Descripción del campo
Keyarea	char	18	not null	primary key		codigo del area de investigacion
codigoOLTP	char	18	not null			
descripcionArea	char	18	not null			nombre del area de investigacion
Keycarrera	char	18	not null	Foreign key(t_dimensioncarreraprofesional)		codigo de carrera profesional

Tabla N° 15 Modelado Físico Tabla T_Dimensión área investigación

T_DIMENSION TIEMPO DIMENSION QUE ALMACENA EL TIEMPO						
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keytiempo	char	18	not null	primary key		
Fecha	char	18	not null			fecha
Keymes	char	18	not null	Foreign key(t_tiempos)		codigo del mes

Tabla N° 16. Modelado Físico Tabla T_Dimensión tiempo

T_DIMENSION DOCENTE DIMENSION QUE ALMACENA LA INFORMACION DEL DOCENTE						
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keydocente	char	18	not null	primary key		codigo del docente
codigoOLTP	char	18	not null			
Nombredocente	char	18	not null			nombre del docente que va a dictar el diplomado
Carreraprofesional	char	18	not null			carrera profesional

Tabla N° 17 Modelado Físico Tabla T_Dimensión docente

T_DIMENSION PRODUCTO DIMENSION QUE ALMACENA LA INFORMACION DEL PRODUCTO						
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
keyproducto	char	18	not null	primary key		codigo del producto
codigoOLTP	char	18	not null			
Descripcionproducto	char	18	not null			nombre del producto que se va a dictar
areainvestigacion	char	18	not null			area de investigacion a la cual pertenece el producto
carreraprofesional	char	18	not null			carrera profesional

Keycarrera	char	18	not null	Foreign key(t_dimensioncarreraprofesional)	codigo de carrera
-------------------	------	----	----------	--	-------------------

Tabla N° 18. Modelado Físico Tabla T_Dimensión producto

T_DIMENSION CARRERA PROFESIONAL		DIMENSION QUE ALMACENA LA INFORMACION DE LA CARRERA PROFESIONAL				
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keycarrera	char	18	not null	primary key		codigo de la carrera
codigoOLTP	char	18	not null			
Descripcioncarrera	char	18	not null			describir la carrera profesional

Tabla N° 19 Modelado Físico Tabla T_Dimensión carrera profesional

T_TIEMPO MES		TIEMPO EN QUE SE REALIZAN LOS DIPLOMADOS				
Campos	Tipo de dato	Tamaño	Nulidad	Restricción	Descripción de la restricción	Descripción del campo
Keymes	char	18	not null	primary key		codigo del mes
Mes	char	18	not null			numero de mes
Trimestre	char	18	not null			numero de trimestre
Semestre	char	18	not null			numero de semestre

Tabla N° 20 Modelado Físico Tabla T_Tiempo mes

Las tablas que han sido presentadas anteriormente representa el nombre de las tablas del datamart con sus respectivos campos, tipos de datos, tamaño del campo, si son nulos o no, si presentan restricciones los campos y su breve descripción de los campos para así saber cuál es el significado que tiene cada uno de los campos en las respectivas tablas.

4.6 DISEÑO Y DESARROLLO DE PRESENTACION DE DATOS

Construcción de ETL de los Datos

Para el proceso de extracción, transformación y carga de los datos se utilizó como herramienta el TOS (TALEND OPEN STUDIO)

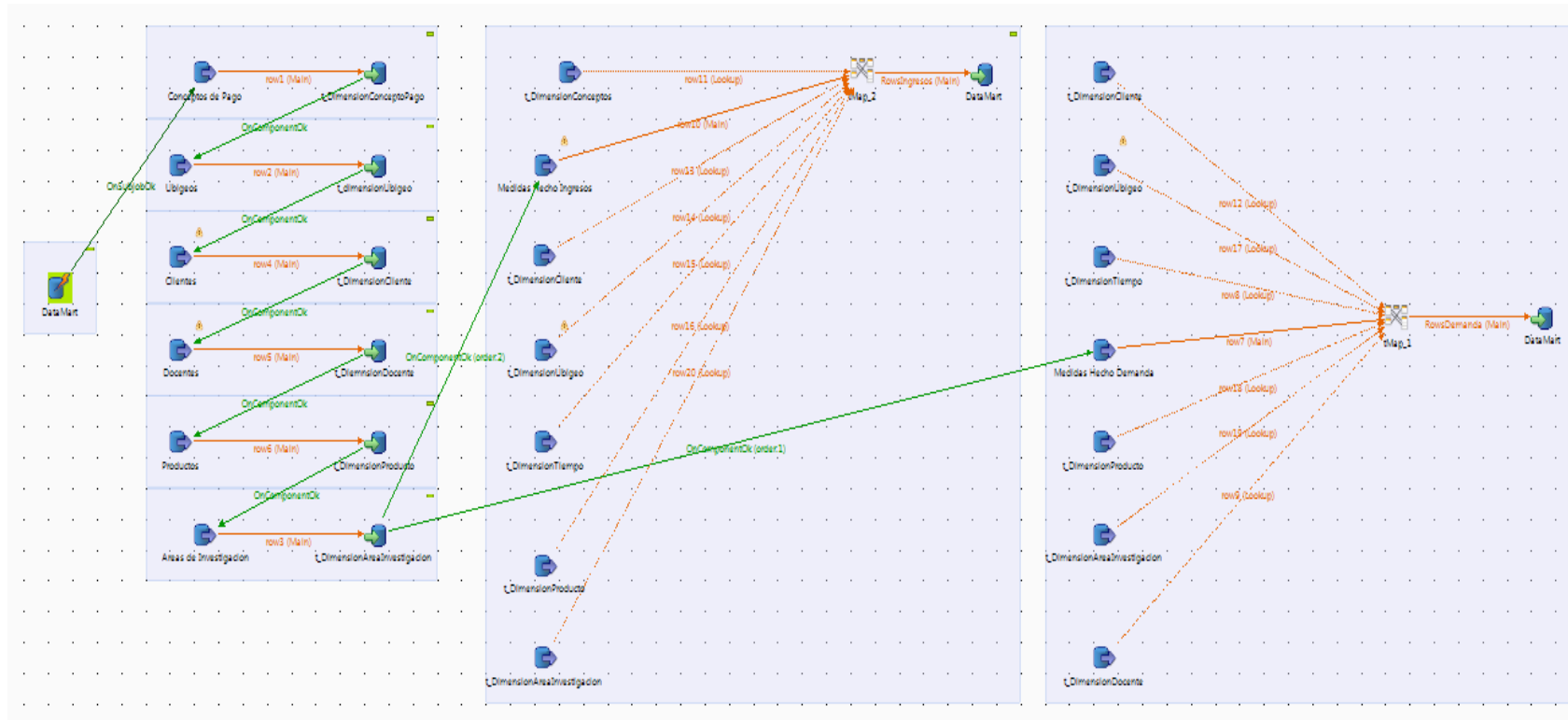


Gráfico N° 07 poblado hechos ingreso y demanda

A través de esta red de extracción, transformación y carga se hace el poblamiento de los datamarts: Ingresos y Demanda

FLUJO DE DATOS

Extracción		Transformación	Destino
Tabla	Campos		t_dimensionconceptopago
Conceptopago	Conceptoid	Conversión de datos	Keyconcepto
	descripcion	Conversión de datos	Codigooltp
			Descripcionconcepto

Tabla N° 21. Flujo de datos de la tabla concepto pago hacia t_dimensión concepto pago

Extracción		Transformación	Destino
Tabla	Campos		t_dimensionubigeo
Ubigeo	Código		keyubigeo
	distrito		codigooltp
	provincia		distrito
	departamento		provincia
			departamento

Tabla N° 22. Flujo de datos de la tabla ubigeo pago hacia t_dimensión ubigeo

Extracción		Transformación	Destino
Tabla	Campos		t_dimensionconcliente
Alumno	alumnoid		keycliente
	apellidos	Concatenación	codigooltp
	nombres		nombrecompleto
	sexo		sexo
País	nombrepais		pais
Ocupación	descripción		ocupacioncliente

Tabla N° 23. Flujo de datos de la tabla alumno pago hacia t_dimensión cliente

Extracción		Transformación	Destino
Tabla	Campos		t_dimensiondocente
docente	docenteid		keydocente
	apellidos	Concatenación	codigooltp
	nombres		nombredocente
carreraprofesional	descripcioncarrera		carerraprofesional

Tabla N° 24. Flujo de datos de la tabla docente pago hacia t_dimensiondocente

Extracción	Transformación	Destino
Tabla	Campos	t_dimensionproducto
Catalogodiplomado	diplomadoid	keyproducto
	descripciondiplomado	codigooltp
Areainvestigacion	descripción	areainvestigacion
carreraprofesional	descripcioncarrera	carreraprofesional

Tabla N° 25. Flujo de datos de la tabla concepto catalogo diplomado hacia t_dimensión producto

Extracción	Transformación	Destino
Tabla	Campos	t_dimensionareainvestigacion
Areainvestigacion	areaid	keyarea
	descripcion	codigooltp
carreraprofesional	descripcioncarrera	nombrearea
	rera	carreraprofesional

Tabla N° 26. Flujo de datos de la tabla áreas investigación hacia t_dimensión área investigación

Extracción	Transformación	Destino
Tabla	Campos	t_demanda
Inscripción	alumnoid	Búsqueda de keycliente en la tabla dimensión t_dimensioncliente
Alumno ediciondiplomado	fechainicio	Búsqueda de keytiempo en la tabla dimensión t_dimensiontiempo
	diplomadoID	Búsqueda de keyProducto en la tabla dimensión t_dimensionProducto
	docenteid	Búsqueda de keyDocente en la tabla dimensión t_dimensionDocente
catalogodiplomado	areaid	Búsqueda de keyarea en la tabla dimensión t_dimensionareainvestigacion

Areasinvestigacion**n**

Ubigeo	codigo	Búsqueda de keyubigeo en la tabla dimensión t_dimensionubigeo	keyubigeo
---------------	--------	---	-----------

Tabla N° 27. Flujo de datos de la tabla inscripción hacia t_demanda

Extracción		Transformación	Destino
Tabla	Campos		t_ingresos
registropago	fecharegistro	Búsqueda de keyTiempo en la tabla dimensión t_dimensiontiempo	keytiempo
detallepago	montocancelado	Conversión de tipo de datos	montototal
obligacionesinscripcion	conceptoid	Búsqueda de keyconceptopago en la tabla dimensión t_dimensionconceptopago	keyconcepto
Inscripción	alumnoid	Búsqueda de keyCliente en la tabla dimensión t_dimensiontiempo	keycliente
Alumno			
ediciondiplomado			
catalogodiplomado	diplomadoid	Búsqueda de keyProducto en la tabla dimensión t_dimensionProducto	keyproducto
	areaid	Búsqueda de keyarea en la tabla dimensión t_dimensionareainvestigacion	keyArea
Ubigeo	codigo	Búsqueda de keyUbigeo en la tabla dimensión t_dimensionubigeo	keyubigeo

Tabla N° 28. Flujo de datos de la tabla registro pago hacia t_ingresos

4.7 DISEÑO DE LA ARQUITECTURA TECNICA.

La arquitectura es la etapa de la metodología enfocada en las herramientas o infraestructura necesaria para dar soporte a los requerimientos del Usuario

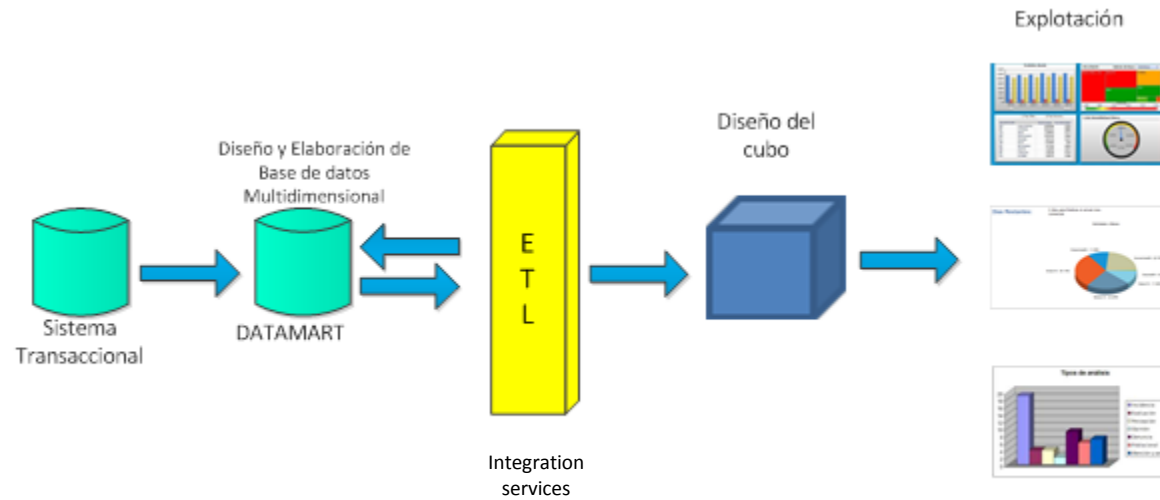


Gráfico N° 08. Arquitectura Técnica de Business Intelligence

4.8 ESPECIFICACIÓN DE APLICACIONES PARA LOS USUARIOS FINALES

Para detallar el proceso de elaboración de las aplicaciones para los usuarios finales primero deberemos conocer los diferentes roles o perfiles de los usuarios que manejarán el Sistema de Soporte a la Toma de Decisiones.

Las herramientas de acceso a la información son el conjunto de aplicaciones que permiten a los usuarios obtener información del Datamart.

La interacción de los usuarios finales con los datos y reportes será a del DSS Web desarrollado en este proyecto.

Para la navegación en el Sistema de Soporte a la Toma de Decisiones Web de Eknowledge group SAC las personas encargadas del acceso a este sistema contarán con un usuario y clave que le permitirá acceder a esta solución de forma fácil en donde podrá encontrar la información de manera rápida y eficiente lo que le permitirá analizar más fácil los reportes de acuerdo a sus necesidades.

Este sistema no solo se podrá acceder desde la misma empresa sino también se podrá acceder a esta aplicación desde cualquier parte del mundo mediante la web.

4.9 Implementación

El sistema de soporte y gestión se implementó con las siguientes tecnologías:

La implementación del poblamiento del datamart con TOS (Talend Open Studio).

La implementación del cubo con una arquitectura ROLAP.

La implementación de la explotación a través de reportes y dashboards para mejorar la toma de decisiones.

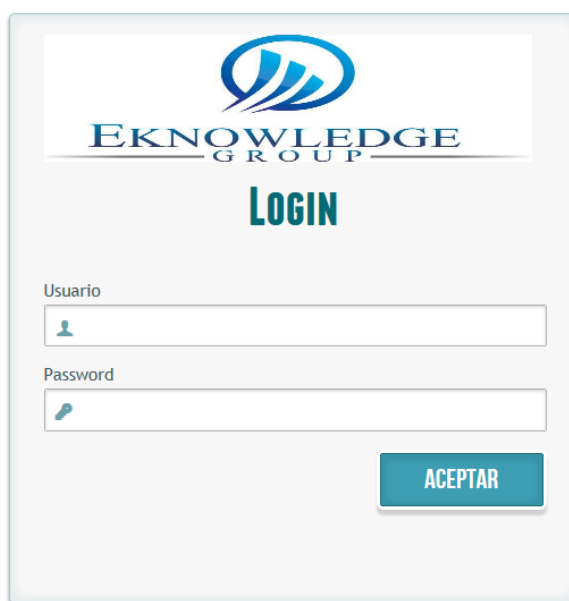
4.10 Mantenimiento

La programación del ETL se va a realizar diariamente.

En lo concerniente al backups eso se encarga el transaccional.

Las operaciones de mantenimiento de datos por ejemplo si se tienen datos de un año como el 2005 se van a eliminar del datamart y cuando los necesiten solo se incluyen.

5. Desarrollo de aplicaciones para usuarios finales



The image shows a login interface for EKNOWLEDGE GROUP. At the top center is the company logo, which consists of a stylized blue wave or leaf shape above the text 'EKNOWLEDGE GROUP'. Below the logo, the word 'LOGIN' is displayed in a large, bold, blue font. Underneath 'LOGIN', there are two input fields. The first is labeled 'Usuario' and contains a small person icon. The second is labeled 'Password' and contains a small key icon. To the right of these fields is a blue button with the word 'ACEPTAR' in white capital letters.

Gráfico N° 09 Acceso a Usuario EknowledgeGroup SAC
Pantalla inicial al Sistema de Gestión y Soporte para la Toma de Decisiones donde se realiza la validación del ingreso a los usuarios del sistema.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

[Analisis Demanda ▶](#) [Analisis Ingresos ▶](#) [Analisis Varios ▶](#) [Cerrar Sesión](#)



Gráfico N° 10 Menú Principal del Sistema de Gestión y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC

El presente menú es la página principal donde el usuario empezará a interactuar con el sistema para poder elaborar sus reportes y dashboards y así poder obtener de forma rápida y eficaz lo solicitado por su superior.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

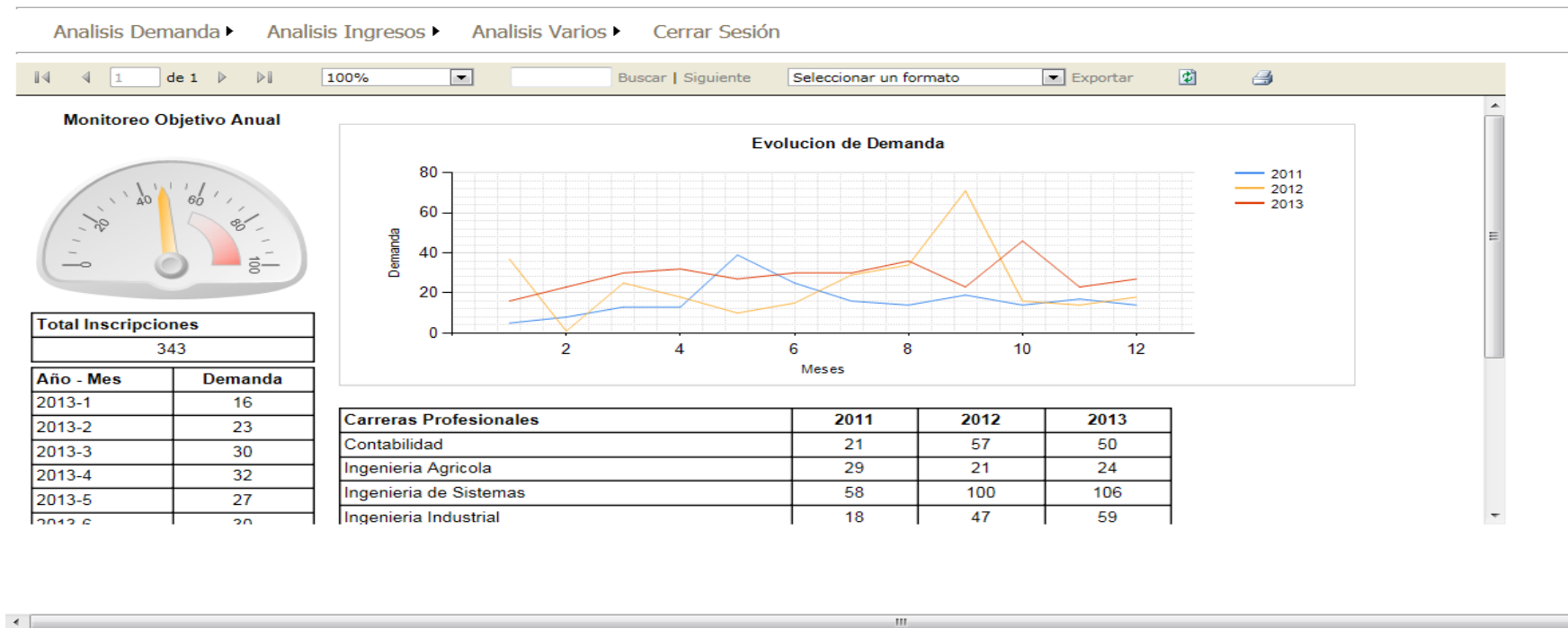


Gráfico N° 11 Dashboard Demanda

Este dashboard muestra la demanda que se ha ido obteniendo con el avance de los años, en los diferentes diplomados que se han ido realizando en las diferentes área de las carreras profesionales, como se puede apreciar en el dashboard vemos que en el 2011 la carrera de ingeniería de sistemas tenía un total de 58 inscritos y en el 2013 aumento a 106 inscritos, así como en la carrera de contabilidad vemos el incremento del año 2011 con 21 inscritos y en el 2013 se incrementó a 50.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

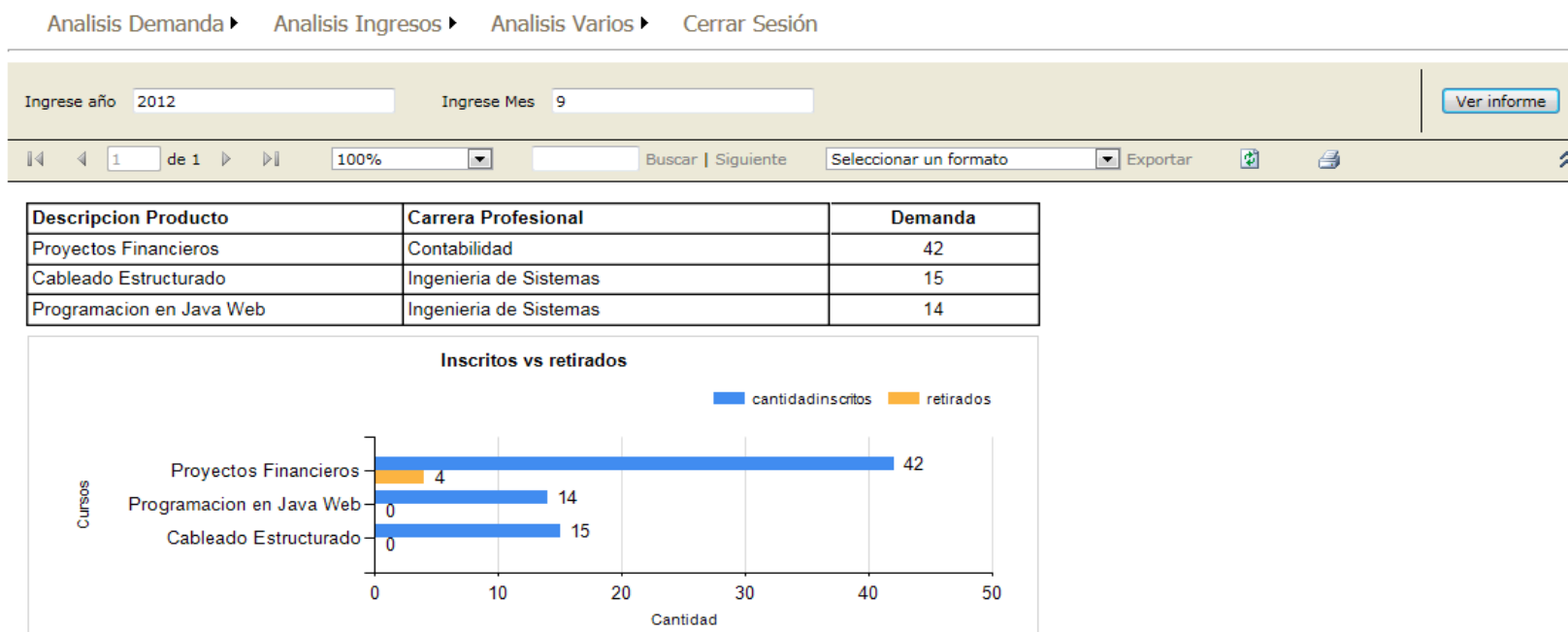


Gráfico N° 12 Reporte de los cursos con la cantidad de inscritos y retiros

Este reporte nos muestra un comparativo entre los cursos que se han dictado y cuantos han optado por retirarse del diplomado por alguna razón. Como se puede apreciar los cursos de java web y de cableado estructurado no presentan retiros lo que indica que tiene aceptación por los clientes de la empresa educativa.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

Analisis Demanda ▶ Analisis Ingresos ▶ Analisis Varios ▶ Cerrar Sesión

CarreraProfesional AreaInvestigacion

1 de 2 100% Buscar | Siguiete Seleccionar un formato Exportar

Nombre Completo Alumno	Descripcion Producto	Cantidad
Abad Abad, Alberto	Configuracion de servidores Linux	2
	Seguridad informaticas	1
Ayala Custodio, Fernando	Seguridad informaticas	1
Balcazar Zapata, Silvia	Configuracion de servidores Linux	1
	Seguridad informaticas	1
Ballena Primo, Gerardo	Configuracion de servidores Linux	1
	Seguridad informaticas	1
Bances Perez, Carolina	Seguridad informaticas	1
Barrantes Carrasco, Catalina Enriqueta	Seguridad informaticas	2
Benavides Nuñez, Dora	Seguridad informaticas	1
Calderon Flores, Alan	Seguridad informaticas	1
Camacho Calle, Pablo	Seguridad informaticas	1
Campos Perez, Isabel	Configuracion de servidores Linux	1
	Seguridad informaticas	2
Cancino Medina, Tatiana	Seguridad informaticas	1
Carbone Zamora, Victoria	Cableado Estructurado	1

Gráfico N° 13 Ocurrencias cliente por área.

Este reporte muestra el número de cantidades que una persona se ha inscrito en los diplomados que son emitidos por la empresa, para así tener un conocimiento de cuáles son los clientes que se pueden tomar en cuenta en un próximo diplomado que realice la empresa educativa.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

Analisis Demanda ▶ Analisis Ingresos ▶ Analisis Varios ▶ Cerrar Sesión

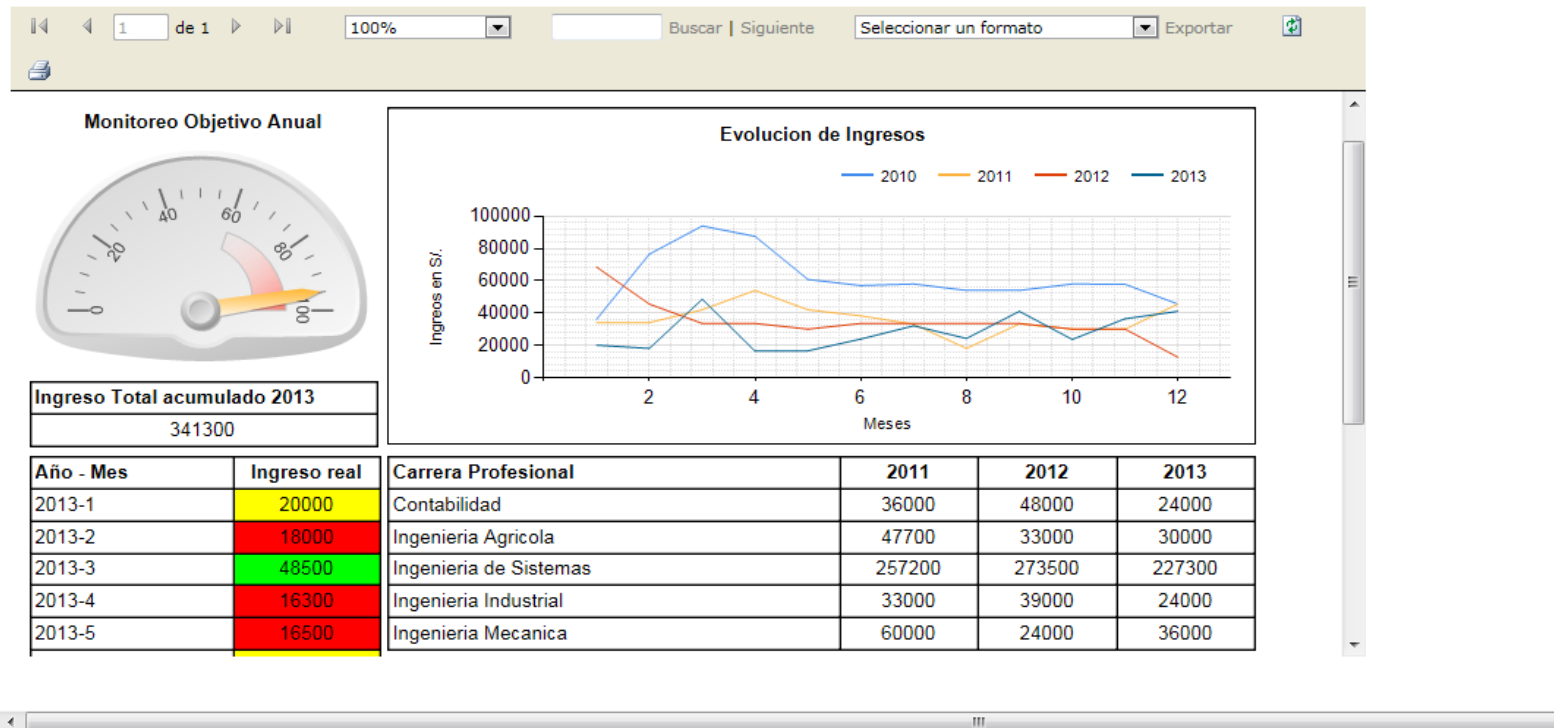


Gráfico N° 14 Dashboard Ingreso

En este dashboard se puede observar en el gráfico los ingresos totales por meses en un determinado año, y por las diferentes carreras profesionales por ejemplo podemos ver que los ingresos mayores en el año 2011 fueron de Ing. Mecánica y los de menos ingresos en el 2013 fueron de contabilidad junto con Ing. Industrial.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

[Analisis Demanda ▶](#)
[Analisis Ingresos ▶](#)
[Analisis Varios ▶](#)
[Cerrar Sesión](#)

Carrera Profesional Ingenieria de Sistemas ▼

1 de 1 | 100% | Buscar | Siguiente | Seleccionar un formato | Exportar

Productos/Año		1	2	3	4	5	6	7	8	9	10	11	12
Configuracion de servidores Linux	2013	10500	3000	6500	3000	2500	2800	3500	3700	9000	4000	5000	4000
	2010	9000	15000	4000	3500	3000	3000	4000	21000	33000	18000	3800	3500
Programacion en Java Web	2011	4000	4000	12000	3000	9000	4200	6000	6000	3500	6000	15000	3000
	2012	8000	3500	3500	3500	12000	3500	3500	3500	3500	9000	3000	3500
	2013	3500	6000	15000	3800	4000	12000	3000	3500	5000	3800	3500	4000
Seguridad informaticas	2010	18000	21000	15000	21000	3800	9000	24000	24000	15000	4000	21000	24000
	2011	24000	24000	24000	24000	3000	4000	3000	9000	24000	24000	15000	3500
	2012	3500	27000	30000	30000	9000	6000	15000	30000	30000	21000	6000	6000
	2013	6000	9000	27000	3500	4000	3000	4500	5000	3000	3800	3900	30000

Gráfico N° 15 Ciclo de vida de un producto

El siguiente reporte permite analizar el ciclo de vida de un diplomado durante los diferentes meses por un determinado año, para así poder saber en qué mes se produce una estacionalidad o un incremento del algún diplomado, el cual servirá saber que diplomado debe dictarse en un determinado mes del año.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

Analisis Demanda ▶ Analisis Ingresos ▶ Analisis Varios ▶ Cerrar Sesión

Seleccione Año: 2013 ▼

Ver informe

1 de 1

100%

Buscar | Siguiete

Seleccionar un formato ▼ Exportar

Buscar texto

Año - Mes	ingreso
2013-1	20000
2013-2	18000
2013-3	48500
2013-4	16300
2013-5	16500
2013-6	23800
2013-7	32000
2013-8	24200
2013-9	41000
2013-10	23600
2013-11	36400
2013-12	41000
	341300

Ingresos Mensuales por año.

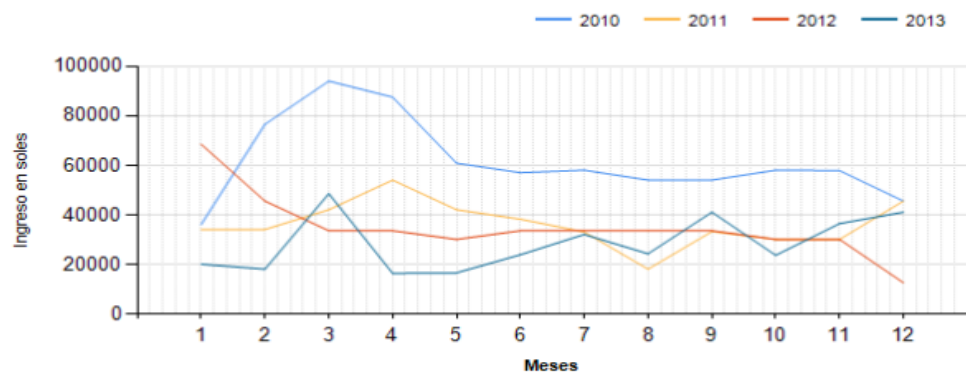


Gráfico N° 16 Evolución Ingreso

El siguiente reporte permite analizar la evolución de los ingresos que ha tenido la empresa ya sea por mes y por un determinado año, lo cual permite ver el progreso que ha obtenido la empresa con el tiempo y así ver una tendencia sobre sus ingresos y planificar que diplomados ofrecer a sus clientes.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

Analisis Demanda ▶ Analisis Ingresos ▶ Analisis Varios ▶ Cerrar Sesión

Ingrese Año: 2013 Ver informe

1 de 1 100% Buscar | Siguiete Seleccionar un formato Exportar

Producto por Areas de Investigacion			Monto Ingreso	
☐ Contabilidad	☐ Formulacion de Estados Financieros	Proyectos Financieros	24000	
☐ Ingenieria Agricola	☐ Mecanica de Suelos	Mermado de Insumos	30000	
☐ Ingenieria de Sistemas	☐ Desarrollo	Programacion en Java Web	67100	
		☐ Redes	Configuracion de servidores Linux	57500
			Seguridad informaticas	102700
☐ Ingenieria Industrial	☐ Tecnologia de Manufactura	Manufacturacion de de Materias Primas	24000	
☐ Ingenieria Mecanica	☐ Mecanica de Fluidos	Estructura y Composicion de Fluidos	36000	

Gráfico N° 17 Ingreso Bruto por Producto

El siguiente reporte nos permite determinar cuáles son las carreras profesionales que han obtenido más ingresos, con su respectiva área a la que pertenece y su diplomado que se ha sido dictado por la empresa. Como se puede apreciar los diplomados con menos entrada son Proyectos financieros y manufacturación de materias primas y el de más ingreso es el diplomado de programación en java web.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.



Gráfico N° 18 Evaluación Docente.

El siguiente reporte nos permite analizar cuáles son los docentes de las diferentes carreras profesionales con mayor aceptación por parte de los clientes a la hora de llevar un diplomado. En este reporte vemos que el docente pertenece a la carrera profesional de Ing. De sistemas el cual presenta un 34.85% de retirados en el diplomado realizado por la empresa.

Sistema de Gestion y Soporte para la Toma de Decisiones EKNOWLEDGEGROUP SAC.

[Analisis Demanda](#) ▶ [Analisis Ingresos](#) ▶ [Analisis Varios](#) ▶ [Cerrar Sesión](#)

Seleccione año

1 de 1 100% Buscar | Siguiente Seleccionar un formato Exportar

anio	mes	ingresos	inscritos
2013	1	20000	16
	2	18000	23
	3	48500	30
	4	16300	32
	5	16500	27
	6	23800	30
	7	32000	30
	8	24200	36
	9	41000	23
	10	23600	46
	11	36400	23
	12	41000	27

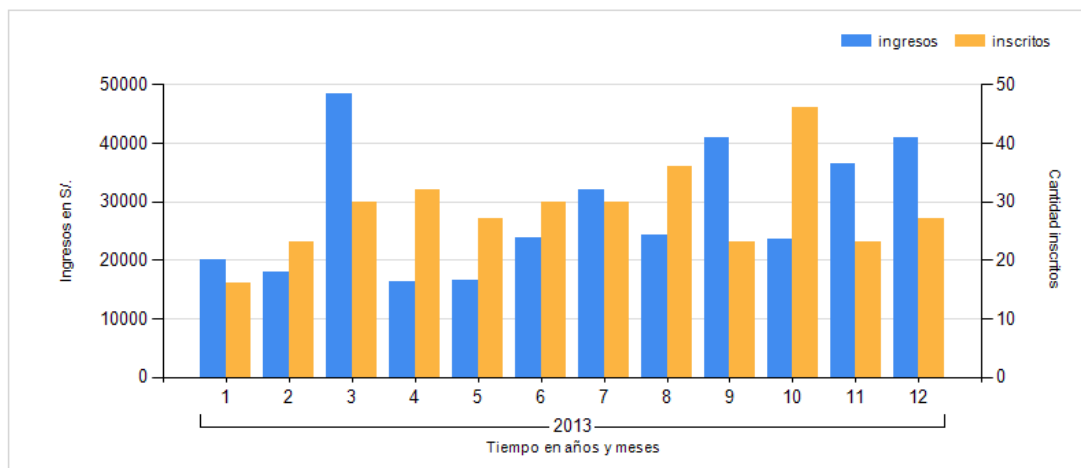


Gráfico N° 19 Reporte Demanda e Ingreso

Este reporte obtenido permite analizar los ingresos de los diferentes meses del año, del mismo modo con la cantidad de inscritos que ha obtenido la empresa a lo largo de los años en los diplomados. Podemos apreciar que en el mes de marzo el ingreso es mayor a los demás, y se puede observar que el mes de enero es donde menos inscritos presenta.

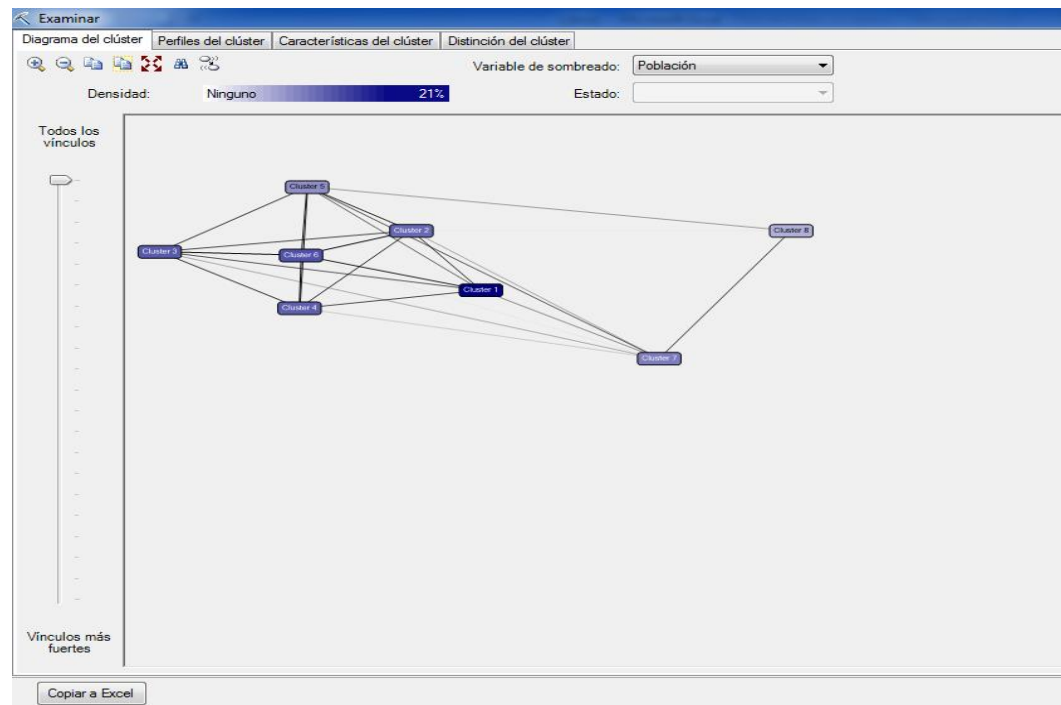


Gráfico N° 20 Elaboración de Clúster

El presente gráfico nos indica cómo se van formando los grupos a través del algoritmo de clúster, el cual se va elaborando de acuerdo a las diferentes variables que se han ingresado para la elaboración de los diferentes perfiles que se van a crear.

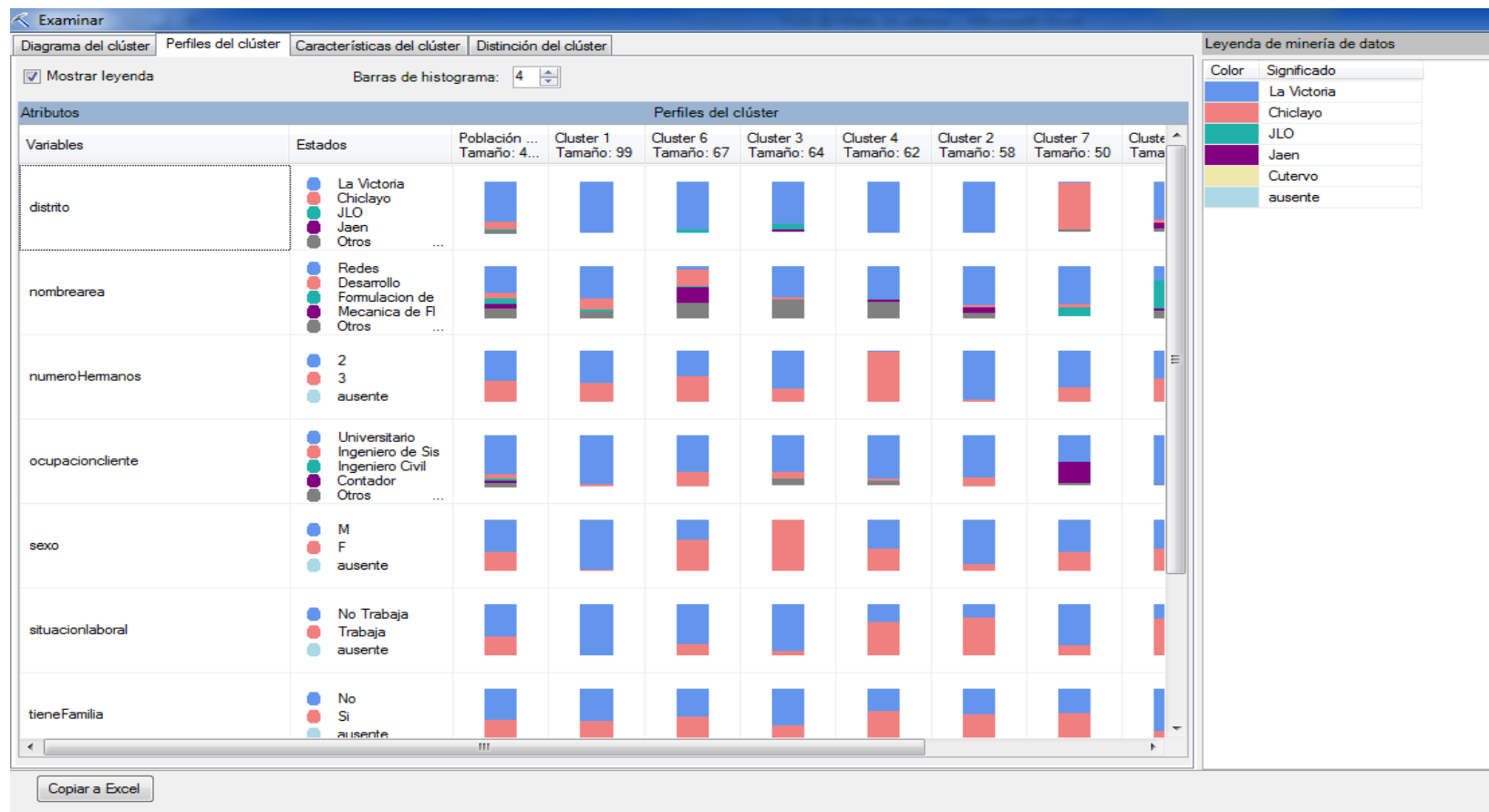


Gráfico N° 21 Perfiles de Clúster

Este diagrama de clúster nos permite saber los perfiles de los distintos grupos que el algoritmo ha elaborado de acuerdo a las diferentes variables que se han ingresado para entrenar al algoritmo. En el primer clúster nos muestra que la mayoría son del distrito de La Victoria, que se han interesado por el curso de redes, presentando un promedio de 2 a 3 hermanos, en su mayoría universitarios de sexo masculino y que no se encuentran trabajando

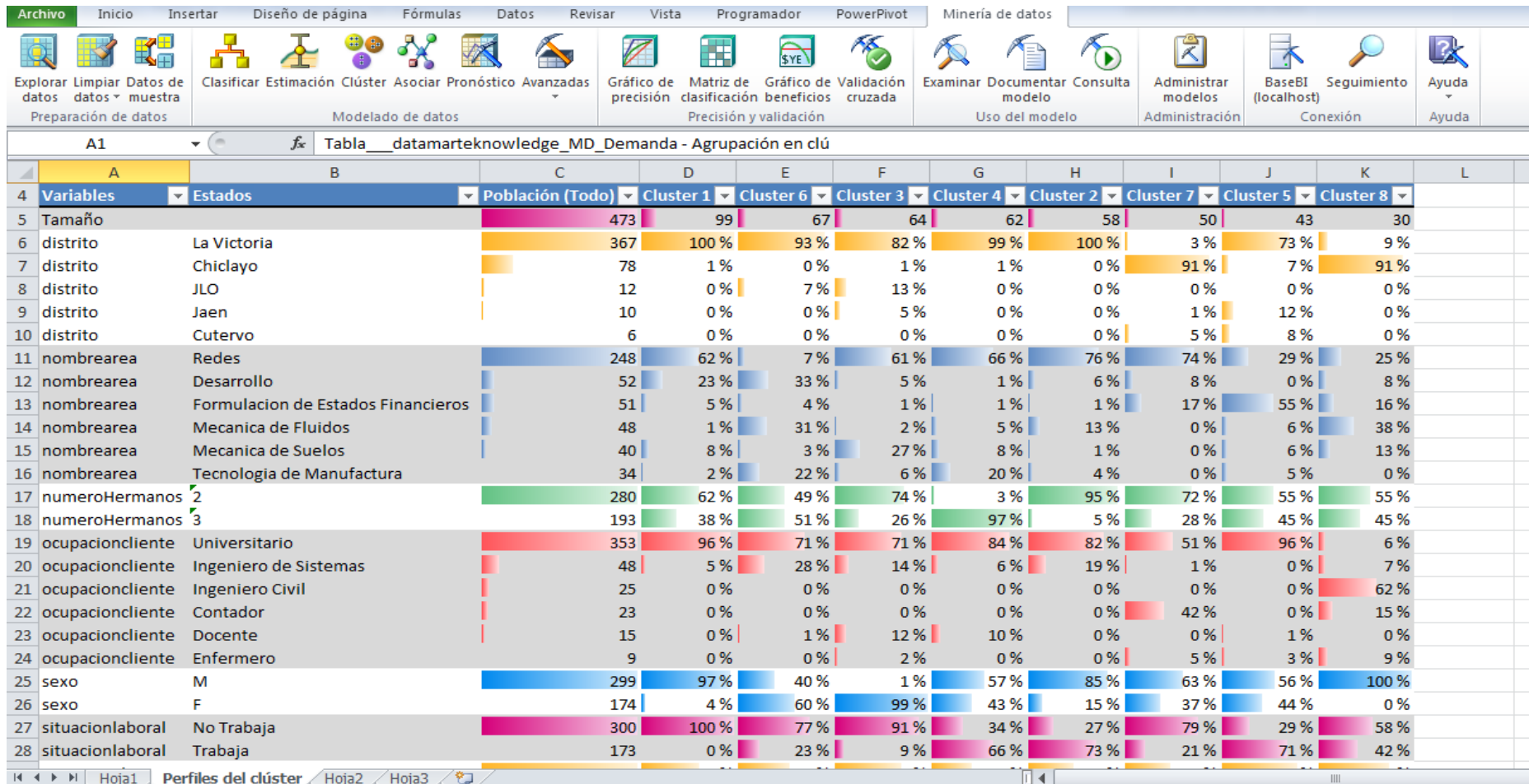


Gráfico N° 22 Perfiles de Clúster en Excell

Este diagrama de clúster nos permite saber los perfiles de los distintos grupos que el algoritmo ha elaborado de acuerdo a las diferentes variables que se han ingresado para entrenar al algoritmo. En el primer clúster nos muestra que la mayoría son del distrito de La Victoria, que se han interesado por el curso de redes, presentando un promedio de 2 a 3 hermanos, en su mayoría universitarios de sexo masculino y que no se encuentran trabajando pero nos muestra las cantidades en porcentaje.

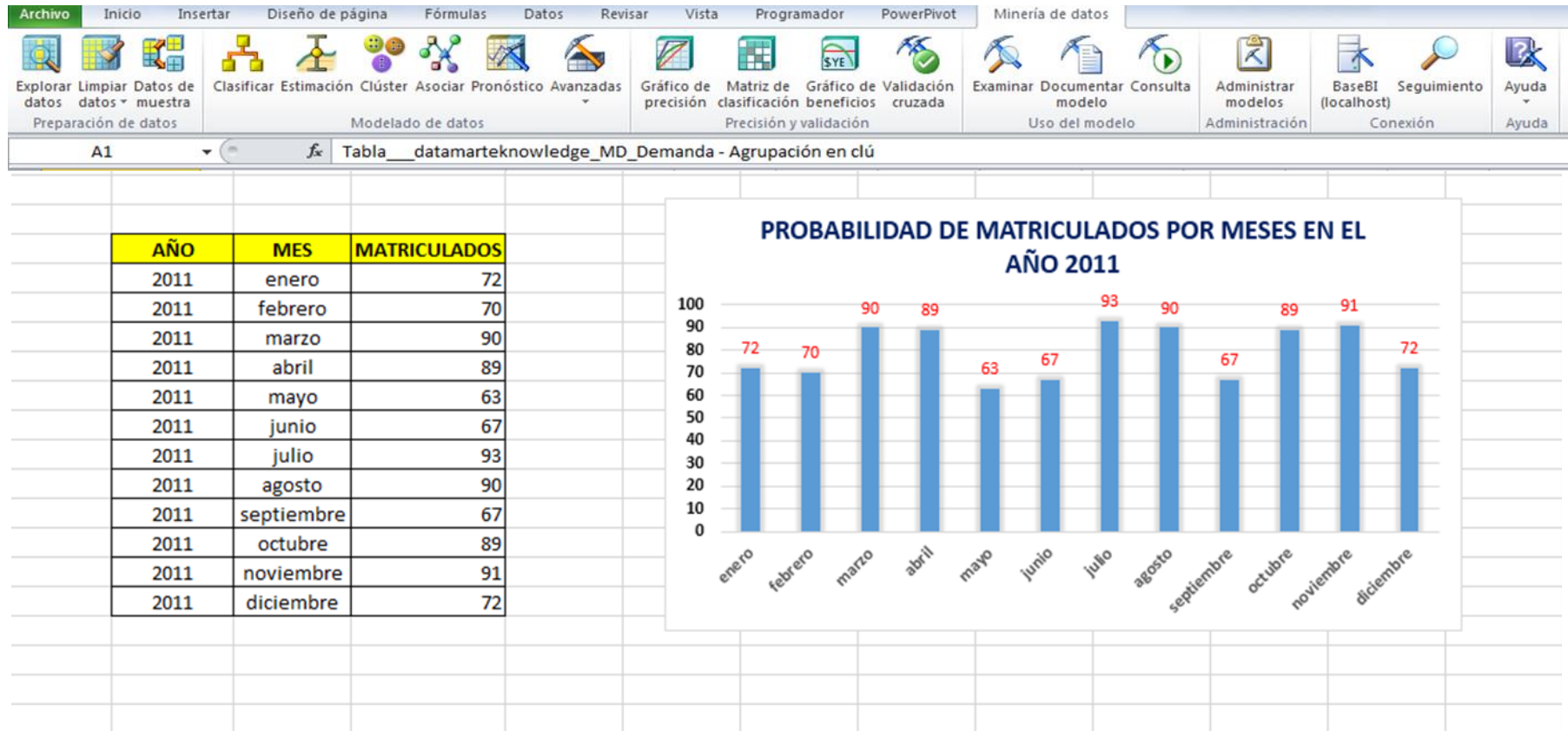


Gráfico N° 23 Probabilidad por el algoritmo de Bayes

Este reporte nos indica la probabilidad de matriculados que puede tener la empresa por meses en un determinado año, y así tener un conocimiento de los ingresos que va a obtener la empresa en el año.

V. DISCUSIÓN

Ejecución del Diseño de contrastaciones Pre test y Post test.

Para contrastar la hipótesis se puso a prueba el sistema que cuenta con la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster para mejorar los procesos analíticos del área comercial de la empresa educativa para la empresa Eknowledge group SAC., para ello se ejecutó una encuesta, dicha encuesta fue aplicada al área gerencial y académico, en la primera etapa con el actual sistema transaccional y la segunda con el sistema propuesto.

Se debe tomar en cuenta que dicha encuesta fue aplicada a 3 ejecutivos de la empresa. Cabe señalar que existen dos tipos de indicadores cuantitativos en los cuales se utilizó como instrumento de medida un cronometro.

Para lo que respecta a indicadores cualitativos se utilizó como instrumento de medida el % a partir de una encuesta antes y después.

Prueba de Hipótesis General

Procedimiento Actual	Duración del Proceso (Pre)	Duración del Proceso (Post)
Extracción de datos	180	25
transformación de datos	120	30
carga de datos	95	25

Tabla N° 29 hipótesis general

Formulación de hipótesis.

H₀ La implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster no mejora los procesos analíticos del área comercial de la empresa educativa.

H₀: $U_{pre} \leq U_{pos}$

H₁ La implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora los procesos analíticos del área comercial de la empresa educativa.

H₁: $U_{pre} > U_{pos}$

Nivel de Significancia

$\alpha=0.05$

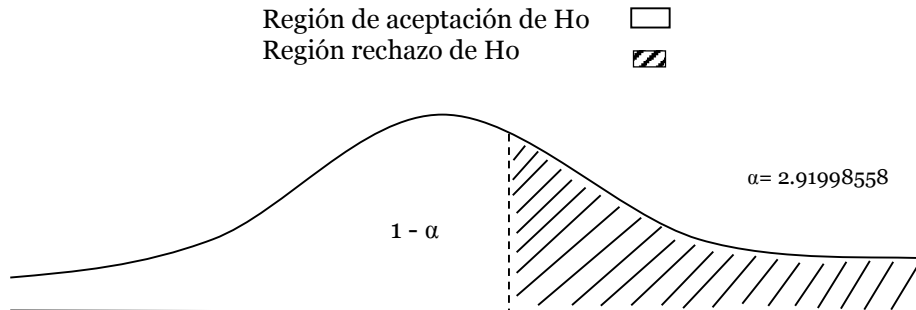
Estadístico de Prueba

$$t_c = \frac{\bar{d}}{sd / \sqrt{n}} \quad t(1 - \alpha, n-1)$$

Región crítica

$$t(1 - \alpha, n-1) = t(1 - 0.05, 3-1)$$

$$= t(0.95, 2) = 2.91998558$$



Valor del Estadístico.

$$tc = \frac{d}{sd / \sqrt{n}}$$

-
- d = diferencia promedio.
- sd= desviación estándar de diferencia promedio.
- n= número de reportes

$$tc = \frac{105}{44.4409721\sqrt{6}} = 2.91998558$$

Prueba t para medias de dos muestras emparejadas

	Duración del Proceso (Pre)	Duración del Proceso (Post)
Media	131.6666667	26.66666667
Varianza	1908.333333	8.333333333
Observaciones	3	3
Coefficiente de correlación de Pearson	-0.23128651	
Diferencia hipotética de las medias	0	
Grados de libertad	2	
Estadístico t	4.092289756	
P(T<=t) una cola	0.0274232	
Valor crítico de t (una cola)	2.91998558	
P(T<=t) dos colas	0.0548464	
Valor crítico de t (dos colas)	4.30265273	

Tabla N° 30. Prueba t de la hipótesis general

Decisión

tc= 4.092289756 pertenece a la Región Crítica
 Por lo tanto se rechaza la Ho

Conclusión

A un 95% de confianza se estima que la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmo de bayes y clúster mejora la toma de decisiones en la empresa educativa E-knowledgegroup SAC.

INDICADORES CUANTITATIVOS

Valores para el indicador: Tiempo en obtener información solicitada por ejecutivos antes y después.

La medición de este indicador se hizo por medio de un cronometro el cual sirvió para medir el promedio de duración que se tiene en cada proceso el análisis realizado en la siguiente tabla de observaciones.

Procedimiento Actual	Duración del Proceso (Pre)	Duración del Proceso (Post)	Diferencia
Reporte ciclo de vida del producto	360	10	350
Reporte de evolución de ingresos	180	10	170
Reporte de ingreso bruto producto	120	12	108
Reporte comparativo demanda ingreso – áreas	300	15	285
Reporte ocurrencias cliente x área	90	8	82
Reporte aceptación docente	90	5	85

Tabla N° 31. Tiempos en obtener información

Prueba de Hipótesis

Formulación de hipótesis.

H₀ Implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster no mejora el tiempo en obtener información del área comercial de la empresa educativa.”

H₀: $U_{pre} \leq U_{pos}$

H₁ Implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora el tiempo en obtener información del área comercial de la empresa educativa.”

H₁: $U_{pre} > U_{pos}$

Nivel de Significancia

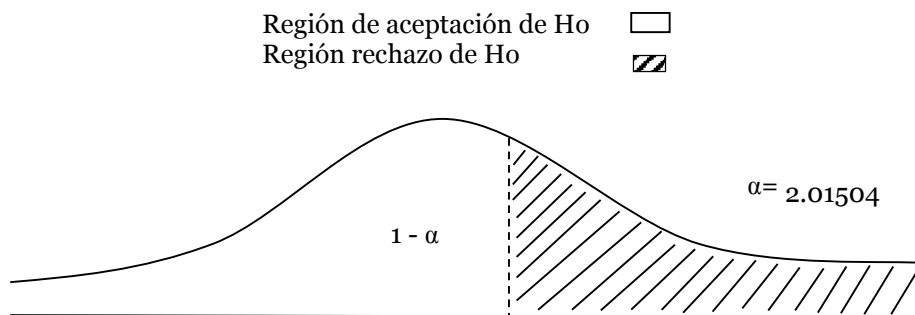
$\alpha=0.05$

Estadístico de Prueba

$$tc = \frac{\bar{d}}{sd/\sqrt{n}} \quad t(1-\alpha, n-1)$$

Región crítica

$$\begin{aligned} t(1-\alpha, n-1) &= t(1-0.05, 6-1) \\ &= t(0.95, 5) = 2.01504 \end{aligned}$$



Valor del Estadístico.

$$tc = \frac{\bar{d}}{sd/\sqrt{n}}$$

\bar{d} = diferencia promedio.
sd= desviación estándar de diferencia promedio.
n= número de reportes

$$\begin{aligned} tc &= \frac{222.6667}{87.8809801/\sqrt{6}} \\ &= 6.206345 \end{aligned}$$

Prueba t para medias de dos muestras emparejadas

	<i>Duración del Proceso (Pre)</i>	<i>Duración del Proceso (Post)</i>
Media	230	7.333333333
Varianza	7800	3.066666667
Observaciones	6	6
Coeficiente de correlación de Pearson	0.2586303	
Diferencia hipotética de las medias	0	
Grados de libertad	5	
Estadístico t	6.206345393	
P(T<=t) una cola	0.000793034	
Valor crítico de t (una cola)	2.015048373	
P(T<=t) dos colas	0.001586068	
Valor crítico de t (dos colas)	2.570581836	

Tabla N° 32. Prueba t del tiempo en obtener información

Decisión

Tc = 6.206345 pertenece a la Región Crítica
 Por lo tanto se rechaza la H₀

Conclusión

A un 95% de confianza se estima que la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmo de bayes y clúster mejora los procesos analíticos del área comercial de la empresa educativa.

Valores para el indicador: Tiempo del ejecutivo en procesar información

La medición de este indicador se hizo por medio de un cronometro el cual sirvió para medir el promedio de duración que se tiene en cada proceso de toma de decisiones por parte de los ejecutivos.

Procedimiento Actual	Duración del Proceso (Pre)	Duración del Proceso (Post)
Reporte ciclo de vida del producto	150	45
Reporte de evolución de ingresos	100	35
Reporte de ingreso bruto producto	60	25
Reporte comparativo demanda ingreso – áreas	100	35
Reporte ocurrencias cliente x área	100	30
Reporte aceptación docente	90	30

Tabla N° 33. Tiempo del ejecutivo en procesar información

Prueba de Hipótesis

Formulación de hipótesis.

H₀ Implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster no mejora el tiempo del ejecutivo en procesar información.

$$H_0: U_{pre} \leq U_{pos}$$

H₁ Implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora el tiempo del ejecutivo en procesar información.

$$H_1: U_{pre} > U_{pos}$$

Nivel de Significancia

$$\alpha = 0.05$$

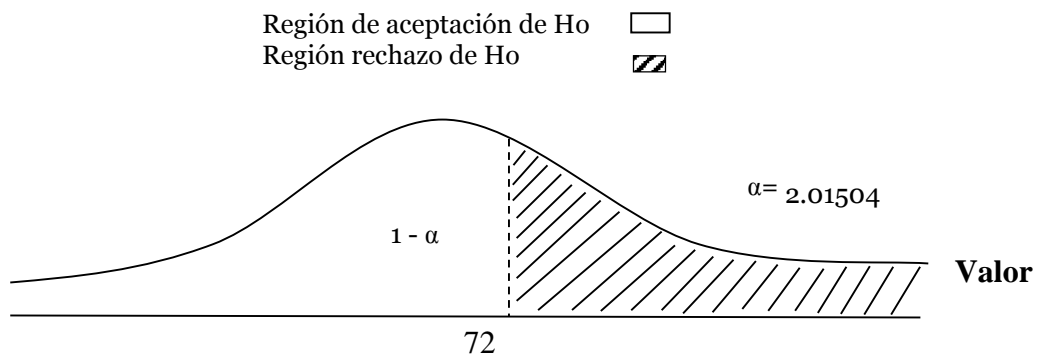
Estadístico de Prueba

$$t_c = \frac{\bar{d}}{sd / \sqrt{n}} \quad t(1 - \alpha, n - 1)$$

Región crítica

$$t(1 - \alpha, n - 1) = t(1 - 0.05, 6 - 1)$$

$$= t(0.95, 5) = 2.01504$$



del Estadístico.

$$tc = \frac{d}{sd / \sqrt{n}}$$

\bar{d} = diferencia promedio.

sd= desviación estándar de diferencia promedio.

n= número de reportes

$$tc = \frac{66.6666667}{22.5092574\sqrt{6}} = 7.2547625$$

Prueba t para medias de dos muestras emparejadas

	<i>Duración del Proceso (Pre)</i>	<i>Duración del Proceso (Post)</i>
Media	100	33.33333333
Varianza	840	46.66666667
Observaciones	6	6
Coefficiente de correlación de Pearson	0.959644917	
Diferencia hipotética de las medias	0	
Grados de libertad	5	
Estadístico t	7.254762501	
P(T<=t) una cola	0.000388666	
Valor crítico de t (una cola)	2.015048373	
P(T<=t) dos colas	0.000777332	
Valor crítico de t (dos colas)	2.570581836	

Tabla N° 34 Prueba t para procesar información

Decisión

tc= 7.2547625 pertenece a la Región Crítica

Por lo tanto se rechaza la H_0

Conclusión

A un 95% de confianza se estima que la implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmo de bayes y clúster mejora la toma de decisiones en la empresa educativa E-knowledgegroup SAC.

Valores para el indicador: Inexactitud de Información

La medición de este indicador se hizo por medio de un conteo el cual sirvió para medir el promedio de errores que presentan los reportes que son emitidos a los ejecutivos.

Procedimiento Actual	Error del Proceso (Pre)	Error del Proceso (Post)
ejecutivo 1	3	0
ejecutivo 2	3	1
ejecutivo 3	4	0

Tabla N° 35 Número de errores de los reportes

Formulación de hipótesis.

H₀ La implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster no mejora la exactitud de la información.

$$\mathbf{H_0: } U_{pre} \leq U_{pos}$$

H₁ La implementación de un sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y clúster mejora la exactitud de la información.

$$\mathbf{H_1: } U_{pre} > U_{pos}$$

Nivel de Significancia

$$\alpha = 0.05$$

Estadístico de Prueba

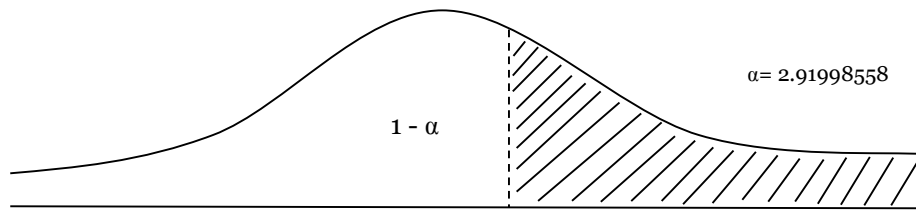
$$t_c = \frac{\bar{d}}{sd / \sqrt{n}} \quad t(1 - \alpha, n-1)$$

Región crítica

$$t(1 - \alpha, n-1) = t(1 - 0.05, 3-1)$$

$$= t(0.95, 2) = 2.91998558$$

Región de aceptación de Ho
 Región rechazo de Ho



Valor del Estadístico.

$$tc = \frac{\bar{d}}{sd / \sqrt{n}}$$

\bar{d} = diferencia promedio.
 sd= desviación estándar de diferencia promedio.
 n= número de ejecutivos.

$$tc = \frac{3}{1\sqrt{6}} = 2.91998558$$

Prueba t para medias de dos muestras emparejadas

	<i>Error del Proceso (Pre)</i>	<i>Error del Proceso (Post)</i>
Media	3.333333333	0.333333333
Varianza	0.333333333	0.333333333
Observaciones	3	3
Coefficiente de correlación de Pearson	-0.5	
Diferencia hipotética de las medias	0	
Grados de libertad	2	
Estadístico t	5.196152423	
P(T<=t) una cola	0.017549359	
Valor crítico de t (una cola)	2.91998558	
P(T<=t) dos colas	0.035098719	
Valor crítico de t (dos colas)	4.30265273	

Tabla N° 36 Prueba t para medir el número de errores de los reportes

Decisión

tc= 5.196152423 pertenece a la Región Crítica
 Por lo tanto se rechaza la Ho

Conclusión

A un 95% de confianza se estima que la implementación de un sistema de gestión y soporte de toma la exactitud de la información mejora la exactitud de la información de los reportes en la empresa educativa E-knowledgegroup SAC.

INDICADORES CUALITATIVOS

Para la medición de los indicadores cualitativos se ha elaborado un cuadro donde se especifica los rangos de valores según el nivel de satisfacción, luego los ejecutivos calificaron a través de una encuesta el nivel de satisfacción según los procedimientos presentados donde se compara el sistema actual versus el sistema propuesto.

Rango de valores	Nivel de satisfacción
[0-2.5]	Muy Malo
<2.5-5]	Malo
<5-7.5]	Bueno
<7.5-10]	Muy Bueno

Tabla N° 37. Escala de medición para indicadores cualitativos

- NIVEL DE SATISFACCIÓN DE LA GERENCIA SOBRE LA INFORMACIÓN SOLICITADO

Indicador	Pregunta	Antes del estímulo		Con estímulo	
		VL	VC	VL	VC
Nivel de Satisfacción Ejecutivo	¿Cuál es su grado de satisfacción con los resultados obtenidos por el sistema?	Malo	3.23	Muy bueno	9.19

Tabla N° 38. Nivel de satisfacción sobre la información solicitada

- NIVEL DE CALIDAD DE INFORMACIÓN SOLICITADA

Indicador	Pregunta	Antes del estímulo		Con estímulo	
		VL	VC	VL	VC
Nivel de Calidad de Información	¿Cómo calificaría la forma en que le son emitidos los reportes por parte de sus trabajadores?	Malo	3.48	Muy Bueno	9.80

Tabla N° 39. Nivel de calidad de la información solicitada

Indicador	VC (M ₁)	VC (M ₂)
Nivel de Satisfacción	3.23	9.19
Nivel de Calidad de la información	3.48	9.8
Total	6.71	18.99

Tabla N° 40. Resumen de Valores de los Indicadores

Hipótesis de prueba

Hipótesis H₀: El sistema actual es mejor que el sistema propuesto.

$$H_0: M_1 - M_2 \geq 0$$

$$H_0: 6.71 - 18.99 = -12.28$$

Hipótesis H₁: El sistema propuesto es mejor que el sistema actual.

$$H_1: M_2 - M_1 > 0$$

$$H_1: 18.99 - 6.71 = 12.28$$

Por lo tanto se concluye que se rechaza H₀ y se acepta H₁ por lo que el sistema propuesto es mejor que el sistema actual, siendo la implementación de la solución de inteligencia de negocios presentado es una alternativa de solución para el problema planteado.

Valores para el indicador cualitativo: Tiempo del ejecutivo en procesar información

Ejecutivo 1

En la encuesta realizada al ejecutivo 1 contestó lo siguiente:

En lo que respecta a la pregunta número uno sobre la información de los reportes que le brindan sus trabajadores, tuvo como respuesta que la

información era regular, el tiempo en cual se demora para la toma de decisiones es de 1 día y medio ya que el tiempo en que se demoran los trabajadores para entregar los reportes es de 2 días lo que ocasiona que se produzca una demora en la toma de decisiones, también se le preguntó sobre la calidad de información en que son presentados los reportes y su respuesta fue de forma regular ya que no se obtenía una información clara y concisa de lo que se pedía puesto que de los reportes que se les pedía por lo menos 3 tenían errores lo que ocasionaba un malestar para el, por lo que los reportes que le eran entregados eran de forma regular por lo tanto su satisfacción en la escala de la pregunta contestó que le era indiferente y para terminar se le preguntó si los reportes que le brindaba su sistema eran analíticos lo cual nos contestó que sí pero que no eran muy claros para poder tomar una adecuada toma de decisiones.

Ejecutivo 2

En la encuesta realizada al ejecutivo 2 contestó lo siguiente:

En lo que respecta a la pregunta número uno sobre la información de los reportes que le brindan sus trabajadores, tuvo como respuesta que la información era buena, el tiempo en cual se demora para la toma de decisiones es de 1 día y medio ya que el tiempo en que se demoran los trabajadores para entregar los reportes es de 2 días lo que ocasiona que se produzca una demora en la toma de decisiones, también se le preguntó sobre la calidad de información en que son presentados los reportes y su respuesta fue de manera buena ya que no se obtenía una información clara y concisa pero veía la forma de cómo entenderlo lo que le ocasionaba pérdida de tiempo en ello sobre el número de errores de los reportes tubo como respuesta de por los menos 3 tenían errores lo que generaba un malestar para el a la hora de tomar alguna decisión con respecto a los diplomados, por lo que los reportes emitidos por los trabajadores eran de forma regular por lo tanto su satisfacción en la escala de la pregunta contestó que le era buena y para terminar se le preguntó si los reportes que le brindaba su sistema eran analíticos lo cual nos contestó que sí pero que no eran muy claros para poder tomar una adecuada toma de decisiones.

Ejecutivo 3

En la encuesta realizada al ejecutivo 3 contestó lo siguiente:

En lo que respecta a la pregunta número uno sobre la información de los reportes que le brindan sus trabajadores, tuvo como respuesta que la información era buena, el tiempo en cual se demora para la toma de decisiones es de 1 día y medio ya que el tiempo en que se demoran los trabajadores para entregar los reportes es de 2 días lo que ocasiona que se produzca una demora en la toma de decisiones, también se le preguntó sobre la calidad de información en que son presentados los reportes y su respuesta fue de manera buena porque los reportes no eran muy entendibles pero veía la forma de cómo entenderlo lo que le ocasionaba pérdida de tiempo en ello sobre el número de errores de los reportes manifestó que siempre habían 3 errores lo cual ya era muy repetitivo por parte de los encargados de realizar esta tarea lo cual llevaba a un malestar para el a la hora de tomar decisiones para la empresa, por lo que los reportes emitidos por los trabajadores eran de forma buena por lo tanto su satisfacción en la escala de la pregunta contestó que le era satisfecho pero hasta cierta forma

ya que algunos si estaban elaborados de forma correcta y por último se le preguntó si los reportes que le brindaba el sistema que tenía la empresa eran analíticos lo cual nos contestó que sí pero que le gustaría que sean mejor elaborados para su correcta toma de decisiones.

Ejecutivos.

En la encuesta realizada a los ejecutivos contestaron lo siguiente:

Después de haber instalado y haber realizado las respectivas pruebas se les hizo la encuesta a los ejecutivos los cuales coincidieron en sus respuestas ya que los reportes que les brindaron los trabajadores fueron muy buenos y para tomar decisiones dijeron que les tomaría solo medio día lo cual los ayuda a programar más rápido las decisiones tomadas el tiempo promedio que se demoran en entregar los reportes fue de 60 minutos, la calidad en que se presentó la información fue muy buena ya que no presentan ningún error los reportes que les fueron entregados y la forma en que se les entrego fue muy buena, dejando al lado los reportes tediosos que les brindaban sus trabajadores, para ello los ejecutivos quedaron muy satisfechos con el grado de satisfacción sobre los reportes y así poder tomar mejor las decisiones con los mejorados reportes analíticos que se les han realizado.

VI. CONCLUSIONES

Después de concluir con la implementación del sistema de soporte de decisiones, se llegó a las siguientes conclusiones:

- ✓ Mejoró la disponibilidad de la información de los procesos analíticos a través de la implementación de los algoritmos de bayes y clúster para la toma de decisiones de la empresa.
- ✓ Se redujo el tiempo promedio en la obtención de información de reportes de resumen de 1140 minutos a 60 min. Esto significa una reducción de 94.75% en el tiempo.
- ✓ Se logró brindar información de calidad de manera rápida, precisa y eficiente.
- ✓ Se logró incrementar el nivel de satisfacción de la gerencia por la información brindada y el fácil manejo del Sistema de soporte de toma de decisiones.

En el Sistema de soporte de decisiones se muestran reportes de gestión, que permite a la gerencia realizar un mejor análisis de lo que está sucediendo en la empresa, además de permitirle crear sus propios reportes y analizar la información de diversas perspectivas.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Anaya Reyna, Luis & Ramírez Bustamante, Franklin, 2009 “Aplicación de business intelligence en la Municipalidad de Chiclayo, utilizando datawarehouse para el soporte en la toma de decisiones urbanísticas y tributarias”. Chiclayo - Perú
- Arsongroup 2009 “Financiera Edificar” <http://www.arsongroup.com/> Perú
- Bismart, 2012 “El Ayuntamiento de Terrassa tomará mejores decisiones gracias a Microsoft Business Intelligence y Bismart”, <http://www.microsoft.com/spain/prensa/noticia.aspx?infoid=2012/02/no15-El-ayuntamiento-de-Terrassa-tomara-mejores-decisiones-gracias-a-Microsoft-BI-y-bismart> España.
- Cohen Karen, Daniel & Asín Lares, Enrique, 2005 “Sistemas de información para los negocios: un enfoque para la toma de decisiones”. México. McGraw-Hill Interamericana.
- Delgado Gonzáles, José & Tarrillo Mondragón Walter, 2008, “Sistema de soporte de decisiones para la gestión operativa en el área comercial de la empresa ferretería Boca EIRL, utilizando software libre”. Chiclayo – Perú
- Evelson Boris, 2008. "Topic Overview: Business Intelligence".
- Fernández Ochoa, Eddy, 2010 “Análisis, diseño e implementación de un datamart de clientes para el área de marketing de una entidad aseguradora”. Lima - Perú
- González Martínez, Marco, 2008 “El Datamart en las pequeñas y medianas empresas como auxiliar en el proceso de toma de decisiones.” Xalapa - Mexico.
- INEI, 1997. “Manual para la construcción de un DataWarehouse”. Lima-Perú.
- Kimball Ralph, 1998. “The Data Warehouse Lifecycle Toolkit”, Wiley.
- Microstrategy, 2006. “Teoría sobre Bussiness Intelligence” 3era Edicion.
- Microsoft. 2012. Algoritmo de clústeres de Microsof. [http://msdn.microsoft.com/es-es/library/ms174879.aspx\(2012\)](http://msdn.microsoft.com/es-es/library/ms174879.aspx(2012))

- Microsoft. 2012. Algoritmo de clústeres Bayes Naive de Microsof. <http://msdn.microsoft.com/es-es/library/ms174806.aspx> (2012)
- Montano Álvarez, Alberto, Mendieta Muñoz, Juan & Reyes Iruegas, Saúl, 2011, Metodología de Business Intelligence para la toma de decisiones efectivas en las empresas pyme. México m
- Núñez Soto, Grace 2010. “Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de finanzas de la Municipalidad Metropolitana de Lima” Lima-Perú.
- Orfelio G. León, 2001 “Tomar Decisiones Difíciles”. España. McGraw-Hill Interamericana.
- Rejado Sedano, Patricia 2008, “Implantación Business Intelligence en Hotel Wellington”, http://www.dharmasigi.com/newsletter/imagenesnews/45_0905/business_intelligence_hotel_wellington.pdf Madrid
- Reyes Reynoso, Mayte & Chalán Santa Maria, Rosa “Prototipo de un daata mart usando herramientas de software libre para apoyar en la toma de decisiones del área de admisión en la clínica Millenium en la ciudad de Chiclayo - Perú” Chiclayo – Perú.
- Silva Lidia & Revello Beatriz, 2008 “Construcción de un sistema de apoyo a la toma de decisiones para el área gerencial del hospital de Clínicas” Montevideo – Uruguay
- Sodexo, 2011 “Implementación de Jaspersoft BI en Sodexo”, http://it-nova.co/productos/inteligencia-de-negocios/caso-de-exito/?utm_source=ZohoCampaign&utm_campaign=Caso%20de%20%E9xito%20-%20Sodexo%20-%20Medell%EDn_2013-01-28_1&utm_medium=email. Colombia
- SSPerú, 2005 “Manufactura: Corporación Miyasato” <http://spperu.com/pdf/microstrategy/Miyasato.pdf>. Perú
- SSPerú, 2005 “MC Procesos Perú y la implementación de soluciones de BI de MicroStrategy”, <http://blog.spasolutioncompany.com/caso-de-exito-mc-procesos-peru-y-la-implementacion-de-soluciones-de-bi-de-microstrategy/> Perú

- Valderrama Mendoza, Santiago, 2010 “Pasos para elaborar proyectos y tesis de investigación científica”. Perú
- Vicente Villardón José “Introducción al Análisis de Clúster”. <http://biplot.usal.es/alumnos/ciencias/2estadistica/multivar/clúster.pdf>. Salamanca
- Villanueva Ojeda, Álvaro, 2011 “Análisis, diseño e implementación de un datawarehouse de soporte de decisiones para un hospital del sistema de salud público” Lima-Perú.
- Vitt, Elizabeth & Luckevich, Michael & Misner & Stacia, 2003 “Business Intelligence: técnicas de análisis para la toma de decisiones estratégicas”, Madrid.

VIII. ANEXOS

Nombre del ingeniero:

Preguntas

1. ¿Qué tipo de decisiones se toman en la organización?
2. ¿cuánto tiempo se demora para tomar esa decisión?
3. Cuenta con dicha información en el momento oportuno para tomar dicha decisión
4. Cuenta con algún sistema transaccional su organización
5. Conoce usted las características de sus clientes a los cuales les interesarían los diplomados que se dictan en su organización
6. Conoce usted cuales son los productos con mayor demanda en las diferentes épocas del año que se dictan en su organización
7. Tiene conocimiento de los ciclos de vida de los productos que ofrece
8. ¿Qué tipo de información le gustaría tener a usted para mejorar su toma de decisión?

Entrevista a Ejecutivos para la Toma de decisiones

Objetivo: Identificar los indicadores

Instrucciones: marcar la respuesta que considere que considere correcta y complete los espacios.

1. Como evaluaría usted la información que le brinda los reportes emitidos por sus trabajadores.

Muy malo (1) Malo (2) Bueno (3) Muy bueno (4)

2. ¿Cuánto tiempo se demora en tomar decisiones con respecto a los diplomados? _____

3. ¿Cuánto tiempo se demoran sus trabajadores para brindarle información sobre los reportes que les son encomendados? _____

4. ¿Cómo evaluaría la calidad de información?

Muy malo (1) Malo (2) Bueno (3) Muy bueno (4)

5. De los reportes que son emitidos cuántos de ellos presentan errores para la toma de sus decisiones _____

6. Cómo calificaría la forma en que les son emitidos los reportes por parte de sus trabajadores.

Muy malo (1) Malo (2) Bueno (3) Muy bueno (4)

7. ¿Cuál es su grado de satisfacción con los resultados obtenidos por el sistema?

Muy insatisfecho (1) Insatisfecho (2) Satisfecho (3) Muy satisfecho (4)

8. El sistema de información genera reportes analíticos para el soporte de la toma de decisiones.

Sí (1) No(0)

Gracias