

“AÑO DEL DIÁLOGO Y LA RECONCILIACIÓN NACIONAL”

UNIVERSIDAD NACIONAL DE HUANCAMELICA

(Creada por Ley N°. 25265)



**FACULTAD DE INGENIERÍA ELECTRÓNICA – SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

TESIS

**“MODELO DE MINERÍA DE DATOS PARA LA MEJORA DEL
SERVICIO ACADÉMICO EN LA INSTITUCIÓN EDUCATIVA
PARTICULAR HARVARD – HUANCAYO”**

**LINEA DE INVESTIGACIÓN
INTELIGENCIA DE NEGOCIOS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

PRESENTADO POR:

BACH. ZORRILLA QUISPE, Nery

BACH. PALOMINO ESPINOZA, María Luz

**HUANCAMELICA – PERÚ
2018**



UNIVERSIDAD NACIONAL DE HUANCAMELICA
(Creada por Ley N° 25265)
FACULTAD DE INGENIERÍA ELECTRÓNICA - SISTEMAS



ACTA DE SUSTENTACIÓN DE TESIS

En el Aula Magna de la Escuela Profesional de Ingeniería de Sistemas, a los **07** días del mes de **Noviembre** del año 2017, a horas **14:00** se reunieron el Jurado Calificador conformado de la siguiente manera:

Presidente : **Dr. Fernando Viterbo SINCHÉ CRISPÍN**
Secretario : **Mg. Ing. Carlos Alcides ALMIDÓN ORTIZ**
Vocal : **Mg. Ing. Everth Manuel RAMOS LAPA**

Ratificados con Resolución N° 382-2017-FIES-UNH del proyecto de investigación (Tesis) Titulado:
"MODELO DE MINERÍA DE DATOS PARA LA MEJORA DEL SERVICIO ACADÉMICO EN LA INSTITUCIÓN EDUCATIVA PARTICULAR HARVARD – HUANCAYO".

Cuyos autores son los graduados:

BACHILLERES: **María Luz PALOMINO ESPINOZA**
Nery ZORRILLA QUISPE

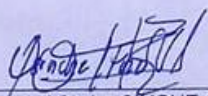
A fin de proceder con la evaluación y calificación de la sustentación del proyecto de investigación, antes citado.

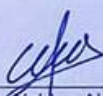
Finalizado la evaluación; se invitó al público presente y a los sustentantes a abandonar el recinto; y luego de una amplia deliberación por parte del jurado, se llegó al siguiente resultado:

APROBADO ☒ **POR MAYORIA**

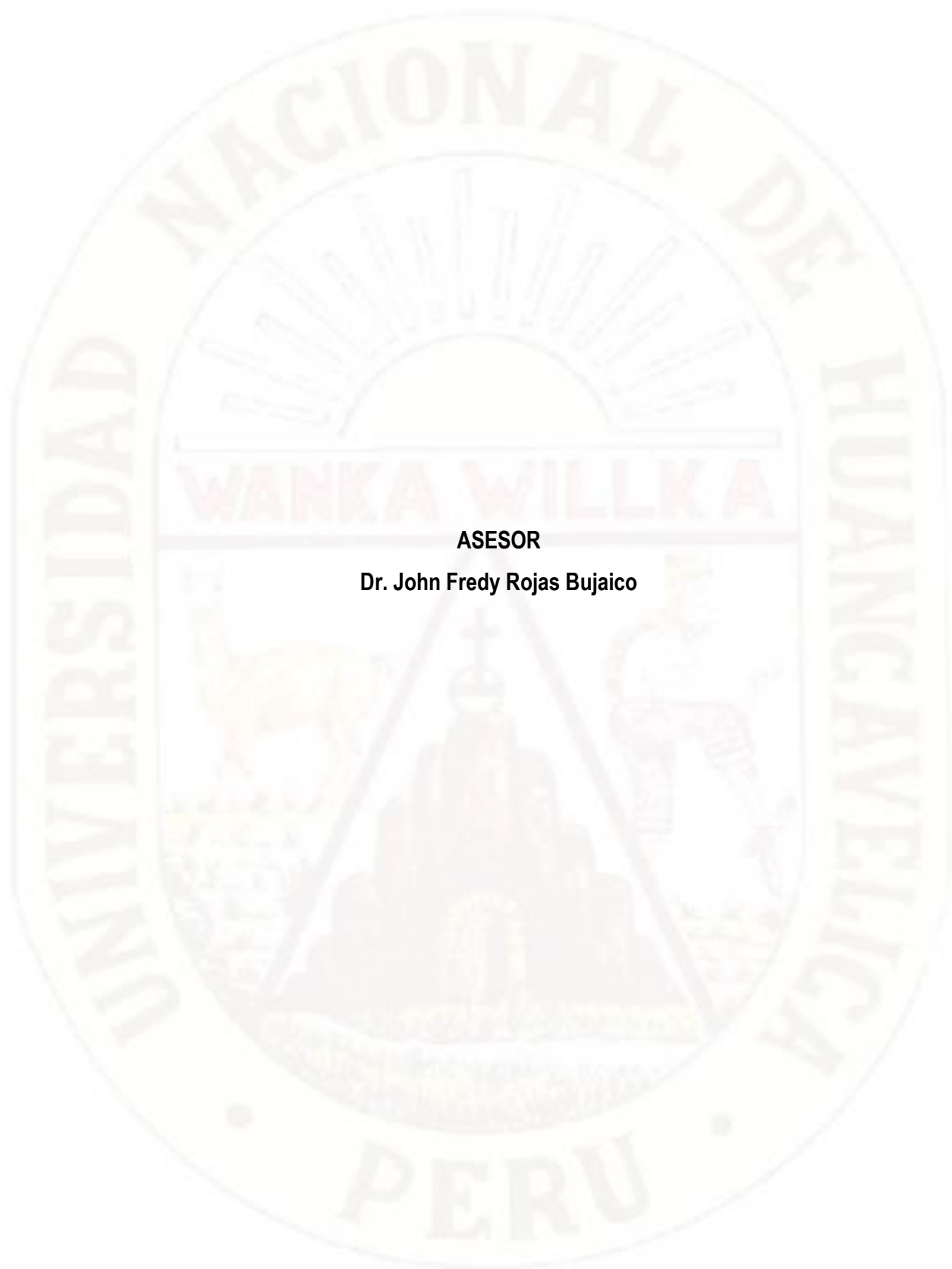
DESAPROBADO ☐

En conformidad a lo actuado firmamos al pie.


Dr. Fernando Viterbo SINCHÉ CRISPÍN
PRESIDENTE

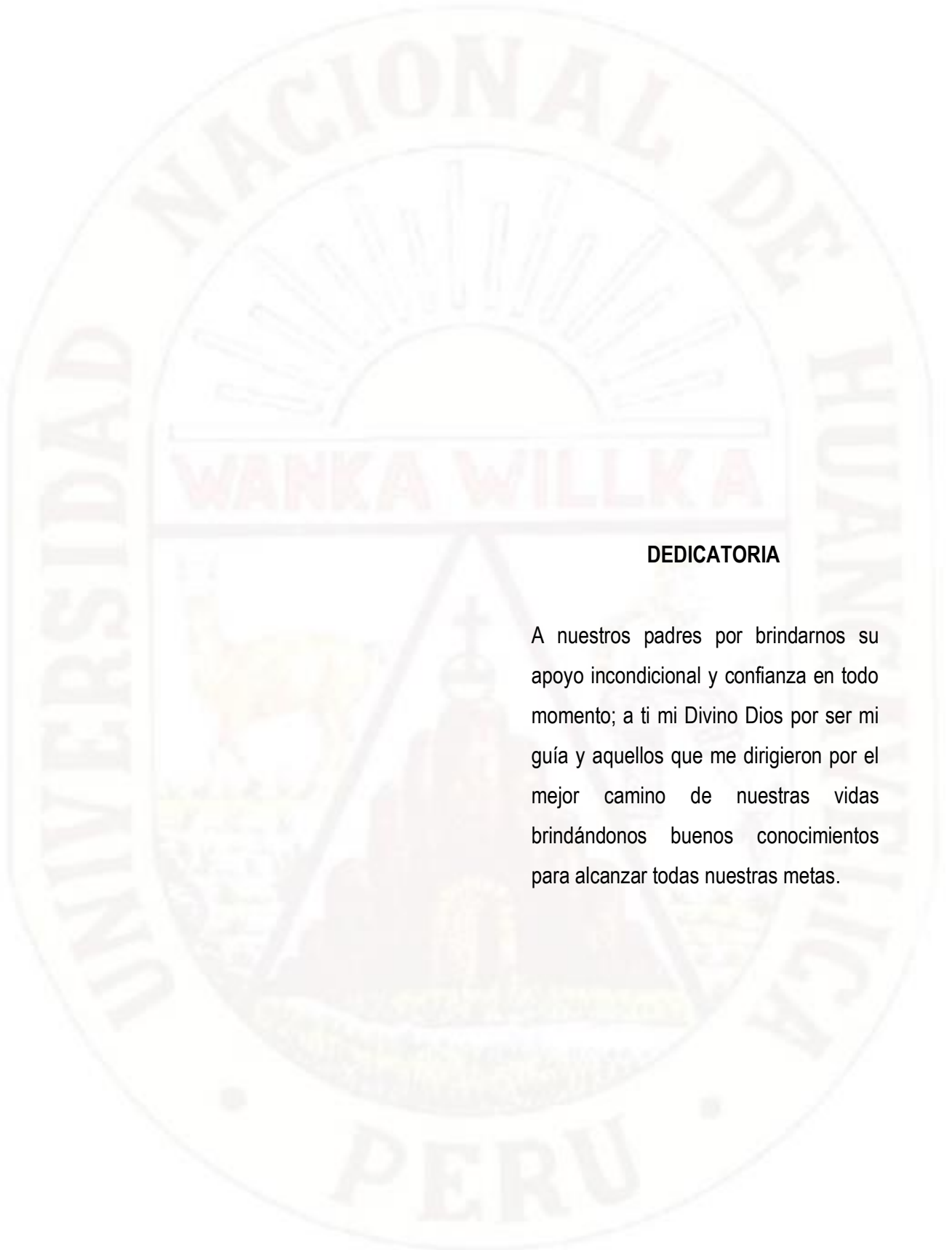

Mg. Carlos Alcides ALMIDÓN ORTIZ
SECRETARIO


Mg. Everth Manuel RAMOS LAPA
VOCAL



ASESOR

Dr. John Fredy Rojas Bujaico



DEDICATORIA

A nuestros padres por brindarnos su apoyo incondicional y confianza en todo momento; a ti mi Divino Dios por ser mi guía y aquellos que me dirigieron por el mejor camino de nuestras vidas brindándonos buenos conocimientos para alcanzar todas nuestras metas.

AGRADECIMIENTO

Agradecemos a nuestro divino Dios por iluminar nuestros caminos y dándonos vida y fuerza para poder subsistir y hacer realidad nuestros logros en este mundo, por iluminar mi camino siempre.

A nuestros queridos padres por sus esfuerzos que están realizando por apoyarnos desinteresadamente en el desarrollo de nuestra meta profesional, por motivarnos a seguir adelante y ser triunfadoras en nuestra sociedad. A ustedes estimados padres les decimos que su apoyo incondicional fue en sí el soporte infinito y que hizo realidad que nosotros culmináramos satisfactoriamente nuestro proyecto profesional.

Igualmente agradezco al Dr. John Fredy Rojas Bujaico, ya que durante la duración de mi proyecto de tesis de investigación me brindó su confianza y contribución de sus conocimientos como profesional los cuales fueron muy valiosos para nosotras.

A la Universidad Nacional de Huancavelica por darnos la oportunidad de ser parte de esta gran familia y especialmente de la familia de la Escuela Profesional de Ingeniería de Sistemas.

Gracias a todos.

ÍNDICE

DEDICATORIA.....	iii
AGRADECIMIENTO	iv
ÍNDICE	v
RESUMEN	viii
ABSTRACT	ix
INTRODUCCIÓN	x
CAPÍTULO I	11
PROBLEMA	11
1.1 Planteamiento del Problema.....	11
1.2 Formulación del Problema	19
1.2.1 Problema General.....	19
1.2.2 Problemas Específicos.....	19
1.3 Objetivos	20
1.3.1 Objetivo General	20
1.3.2 Objetivos Específicos.....	20
1.4 Justificación	20
1.4.1 Justificación Teórica:.....	20
1.4.2 Justificación Metodológica	20
1.4.3 Justificación Práctica	20
CAPITULO II	22
MARCO TEÓRICO	22
2.1 Antecedentes.....	22
2.2 Bases Teóricas.....	24
2.2.1 Inteligencia de Negocio	24
2.2.2 Educación	27
2.2.3 Calidad	28
2.2.4 Pilares de la Calidad Total	30
2.2.5 Fundamentos de la Minería de Datos	35
2.3 Hipótesis	47
2.3.1 Hipótesis General	47
2.3.2 Hipótesis Específicas	47

2.4 Definición de Términos	47
2.5 Identificación de Variables	48
2.5.1 Variable Independiente	48
2.5.2 Variable Dependiente	48
2.6 Definición Operativa de Variables e Indicadores	49
CAPÍTULO III	50
METODOLOGÍA DE LA INVESTIGACIÓN	50
3.1 Ámbito de estudio	50
3.2 Tipo de Investigación	50
3.3 Nivel de Investigación	50
3.4 Método de Investigación	50
3.5 Diseño de la Investigación	51
3.6 Población, Muestra y Muestreo	51
3.7 Técnicas e Instrumentos de Recolección de Datos	52
3.7.1 Técnicas	52
3.7.2 Instrumentos	52
3.8 Procedimiento de Recolección de Datos	52
3.9 Técnicas de Procesamiento y Análisis de Datos	52
CAPITULO IV	53
RESULTADOS	53
4.1 Presentación de Resultados	53
4.1.1 Variable Independiente “Modelo de Minería de Datos”	54
4.1.2 Variable Dependiente “Servicio Académico”	59
4.2 Discusión	68
CONCLUSIONES	70
RECOMENDACIONES	71
REFERENCIA BIBLIOGRÁFICA	72
ANEXOS	73
Anexo N° 01. Matriz de Consistencia	74
Anexo N° 02 Cuestionario Percepción del Servicio Académico	76
Anexo N° 03. Modelo de Minería de Datos	77

ÍNDICE DE GRÁFICOS

Gráfico N° 01: Datos recolectados correspondiente al nivel primario año 2013	12
Gráfico N° 02: Datos recolectados correspondiente al nivel primario año 2014	13
Gráfico N° 03: Datos recolectados correspondiente al nivel primario año 2015	14
Gráfico N° 04: Comparación de datos recolectados años 2013 - 2015	14
Gráfico N° 05: Datos recolectados correspondiente al nivel secundario año 2013.....	15
Gráfico N° 06: Datos recolectados correspondiente al nivel secundario año 2014.....	16
Gráfico N° 07: Datos recolectados correspondiente al nivel secundario año 2015.....	16
Gráfico N° 08: Comparación de datos recolectados años 2013 - 2015	17
Gráfico N° 09: Comparación de datos recolectados años 2013 – 2015	18
Gráfico N° 10: Comparación de datos recolectados años 2013 - 2015	18
Gráfico N° 11 Jerarquía del conocimiento.....	26
Gráfico N° 12 Modelo de calidad educativa	31
Gráfico N° 13 Modelo del sistema de gestión de calidad basado en el ISO 9001	33
Gráfico N° 14 muestra el Ciclo de mejora continua de Deming.....	34
Gráfico N° 15 Etapas de la toma de decisiones	39
Gráfico N° 16 ¿Cuáles son los principales beneficios que su compañía consiguió con la inversión en BI?	42
Gráfico N° 17 Ciclo de vida del Datawarehouse.....	43
Gráfico N° 18 Volatilidad de información.....	44
Gráfico N° 19 Estructura básica de un DataWarehouse.	45
Gráfico N° 20 Proceso ELT (Extracción, Transformación y Carga).....	46

ÍNDICE DE TABLAS

Tabla N° 01.- Operacionalización de variables	49
Tabla N° 02. Resultados obtenidos: Modelo de Minería de datos-Condiciones básicas de infraestructura	60
Tabla N° 03. Modelo de Minería de datos-Condiciones básicas de infraestructura	61
Tabla N° 04. Resultados obtenidos Modelo de Minería de datos - Proceso enseñanza aprendizaje.....	64
Tabla N° 05. Resumen Modelo de Minería de datos - Proceso enseñanza aprendizaje. ..	65

RESUMEN

En la presente tesis titulada “Modelo de minería de datos para la mejora del Servicio Académico en la Institución Educativa Particular Harvard – Huancayo” el problema de investigación fue, ¿ Podría un modelo de minería de datos influir en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo? y el objetivo general fue; Determinar si un modelo de minería de datos podría influir en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo, así también se planteó la hipótesis: Un modelo de minería de datos si influye en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo. El tipo de estudio de la presente investigación es de carácter aplicativa porque se utiliza la minería de datos en el diseño de un modelo del mismo nombre con la finalidad de medir su influencia en la mejora del servicio académico, el nivel de investigación es descriptivo – explicativo. El diseño de investigación fue GE: O1 X...O2, con Pre y Post prueba. Se llegaron a las siguientes conclusiones; ¹. El desarrollo de la implementación de un modelo de minería de datos permitió determinar la herramienta adecuada para mejorar el servicio académico de la Institución Educativa Harvard. ² De la investigación se concluye que un modelo de minería de datos SI influye en la mejora del Servicio Académico en la Institución Educativa Particular Harvard-Huancayo, esto se confirma al haber comprobado que el modelo de minería de datos influye en la dimensión de “condiciones básicas de Infraestructura” y en el “proceso de enseñanza aprendizaje”. ³ Un modelo de minería de datos SI influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo. Se evidencia al corroborarlo que 561 respuestas que representan el 93% a la dimensión de SI de la post prueba respecto de 97 respuestas que representa el 16% a la dimensión de SI de la pre prueba y al observar el valor de la prueba de la Z, en donde $Z_c = 18.095 > Z_t = 1.645$ y $H_0 \in RR$, confirmamos la influencia del Modelo de Minería de datos. ⁴ Se afirma que un modelo de minería de datos SI influyen la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo. Se evidencia al corroborar que 719 respuestas que representan el 96% a la dimensión de SI de la post prueba respecto de 55 respuestas que representa el 8% de SI de la pre prueba y al observar el valor de la prueba de la Z, en donde $Z_c = 18.095 > Z_t = 1.645$ y $H_0 \in RR$, confirmamos la influencia del Modelo de Minería de datos.

Palabras Clave: Modelo, Minería de Datos, Servicio Académico.

ABSTRACT

In this thesis entitled "Data Mining Model for the Improvement of the Academic Service at the Harvard - Huancayo Private Educational Institution" the research problem was, Could a data mining model influence the improvement of the academic service in the Institution Private Education Harvard-Huancayo? and the general objective was Determine if a data mining model could influence the improvement of the academic service at the Harvard-Huancayo Private Educational Institution, as well as the hypothesis: A data mining model if it influences the improvement of the academic service in the Educational Institution Private Harvard-Huancayo. The type of study of this research is of an applicative nature because data mining is used in the design of a model of the same name in order to measure its influence on the improvement of academic service, the level of research is descriptive - explanatory. The research design was GE: O1 X...O2, with Pre and Posttest. The following conclusions were reached ¹. The development of the implementation of a data mining model allowed to determine the appropriate tool to improve the academic service of the Harvard Educational Institution ². From the research it is concluded that a SI data mining model influences the improvement of the Academic Service at the Harvard-Huancayo Private Educational Institution, this is confirmed by having verified that the data mining model influences the dimensions of "basic conditions" of Infrastructure "and in the" teaching-learning process " ³. A SI data mining model influences the improvement of the basic Infrastructure conditions at the Harvard-Huancayo Private Educational Institution. It is evidenced by corroborating that 561 responses representing 93% to the SI dimension of the post test with respect to 97 responses representing 16% to the SI dimension of the pretest and observing the value of the Z test. , where $ZC = 18.095 > Zt = 1.645$ and HoRR, we confirm the influence of the Data Mining Model. ⁴. It is affirmed that a SI data mining model influences the improvement of the teaching-learning process in the Harvard-Huancayo Private Educational Institution. It is evidenced by corroborating that 719 responses representing 96% to the SI dimension of the post test with respect to 55 responses representing the 8% SI of the pretest and observing the value of the Z test, where $ZC = 18.095 > Zt = 1.645$ and HoRR, we confirm the influence of the Data Mining Model.

Keywords: Model, Data Mining, Academic Service.

INTRODUCCIÓN

El presente estudio titulado “Modelo de minería de datos para la mejora del Servicio Académico en la Institución Educativa Particular Harvard – Huancayo” responde a la necesidad de monitorear el servicio académico. Con la finalidad de diseñar y proponer un Modelo de minería de datos para influir en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo. El desarrollo de la investigación se encuentra compuesto por cuatro Capítulos: El capítulo I, trata sobre el Problema, en el cual se detalla el Planteamiento del problema, la formulación del problema, el planteamiento de los objetivos general y específicos, como también la justificación. El capítulo II, trata sobre el marco teórico, en este capítulo presentamos los antecedentes del estudio, las bases teóricas, la hipótesis y las variables de estudio. El capítulo III, se refiere a la Metodología de la Investigación, en este capítulo presentamos el ámbito de estudio, el tipo de investigación, el nivel de investigación, el método, el diseño de nuestra investigación, la población, muestra del estudio, las técnicas y procedimientos de recolección de datos y el procesamiento de datos correspondiente. Y en el capítulo IV, presentamos los resultados, el cual trata sobre la presentación y discusión e interpretación de resultados. Finalmente, se declaran las Conclusiones y Recomendaciones, Referencias Bibliográficas y los Anexos.

LAS AUTORAS

CAPÍTULO I

PROBLEMA

1.1 Planteamiento del Problema

La existencia de instituciones educativas en nuestra ciudad se ha incrementado hoy en día a consecuencia del crecimiento de nuestra población por lo que el surgimiento de diferentes instituciones particulares se han visto necesarias, dando como resultado una competencia entre instituciones ya sean públicas y/o particulares; esto hace que las instituciones en especial las instituciones particulares sientan la necesidad de tomar decisiones acertadas para mantener su respectivo posicionamiento en el mercado educativo, pero esto no se hace tan fácil, debido a que no se tiene información relevante que sirva de base para su respectiva toma de decisiones. Los datos almacenados por las diferentes instituciones educativas están dadas a través de diferentes herramientas o medios físicos, pero no lo tienen en forma ordenada, integrada ni clasificada dando como resultado una toma de decisiones tal cual les parezca adecuada.

Caso Colegio Harvard

El Colegio Harvard se encuentra ubicada en la provincia de Huancayo, distrito de El Tambo, siendo éste el local adecuado a las necesidades y exigencias de la tarea pedagógica. Durante su trayectoria, la institución educativa Harvard ha demostrado desarrollo y crecimiento agigantado en cuanto a su posición y obtención de diferentes lauros.

El grado de referencia por parte de los padres de familia han hecho posible la realización de que esta Institución sea una de las más reconocidas en la región centro, pero frente a la competencia actual que tiene se ve en peligro de disminuir este grado de preferencia ya que las nuevas instituciones que se van creando

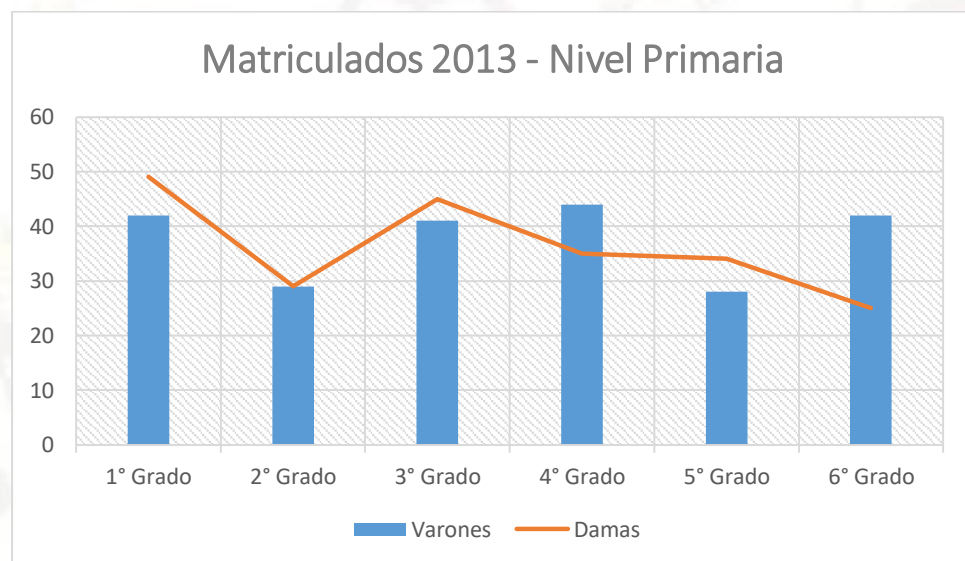
cuentan con tecnología de punta, así como métodos y técnicas de enseñanza-aprendizaje.

Es estudio de investigación que planteamos necesariamente necesita de una base de datos para su implantación, por lo que es necesario coger los datos enmarcados en los 3 últimos años concluidos tanto en el nivel primario y secundario partiendo del año 2013 hasta el 2015.

El orden establecido para la recolección de la información es de la siguiente manera: cantidad de alumnos por grados del nivel primario durante los 3 últimos años, consolidados de alumnos de educación primaria, luego de la misma manera se hace con los del nivel secundario para que finalmente mostrar los alumnos matriculados tanto del nivel primario como del nivel secundario durante los 3 últimos años.

Para el primer año (2013) se ha recolectado los siguientes datos para el nivel de educación primaria, habiendo acogida y posicionamiento dentro de la comunidad educativa, tal como se puede apreciar en el gráfico siguiente:

Gráfico N° 01: Datos recolectados correspondiente al nivel primario año 2013



	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado	6° Grado
Varones	42	29	41	44	28	42
Damas	49	29	45	35	34	25

Fuente: Registros de la Institución Educativa Harvard

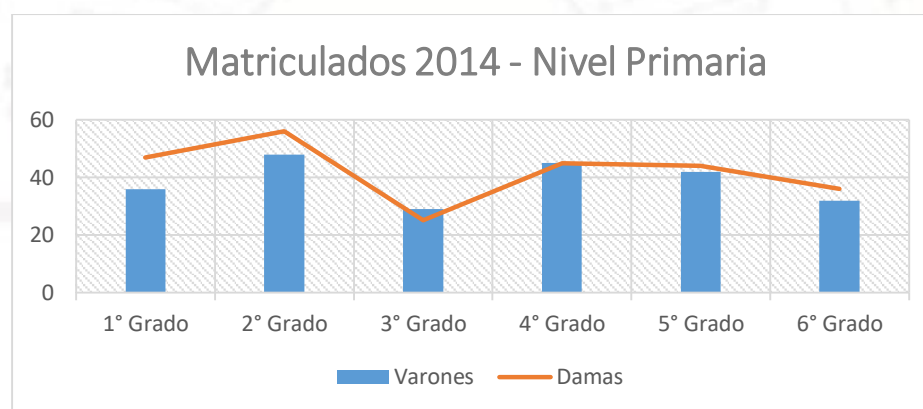
Elaboración: Propia

En el cuadro anterior tenemos los datos distribuidos por cada grado (6 grados) y clasificados por sexo, se puede notar que la cantidad de estudiantes se está

manteniendo en un promedio de 37 alumnos de un total de 443 alumnos. Al contar con un número considerable de alumnos, el manejo de los datos se torna difícil ya que cada estudiante cuenta con diferentes archivos por lo que se tiene que procesar datos para su respectiva toma de decisiones.

Para el 2014, la institución fue mejorando en cuanto a la cantidad de estudiantes matriculados en el colegio, se piensa que dicho incremento se debe a la satisfacción por parte de los padres de familia hacia la institución; los datos recolectados para el año 2014 se muestra en siguiente gráfico:

Gráfico N° 02: Datos recolectados correspondiente al nivel primario año 2014

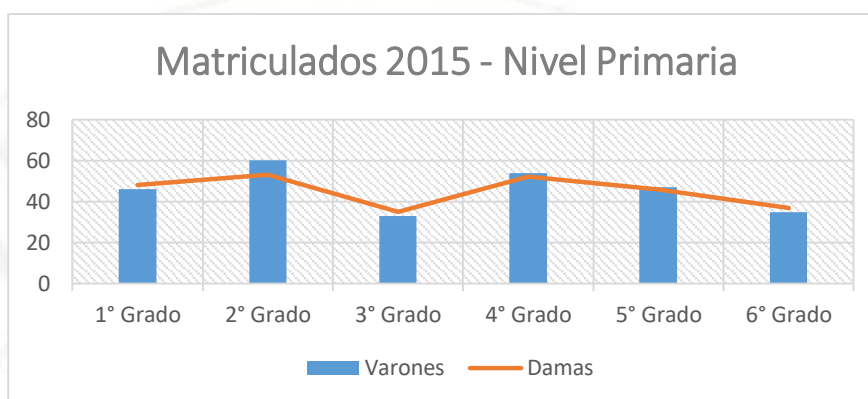


	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado	6° Grado
Varones	36	48	29	45	42	32
Damas	47	56	25	45	44	36

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

Si queremos comparar con los datos recolectados en el año 2013, se puede notar un incremento con referencia al año 2014, también si comparamos los promedios, el promedio actual es de 40 alumnos por cada grado. En el siguiente gráfico podemos visualizar el aumento de alumnos para el año 2015.

Gráfico N° 03: Datos recolectados correspondiente al nivel primario año 2015



	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado	6° Grado
Varones	46	60	33	54	47	35
Damas	48	53	35	52	46	37

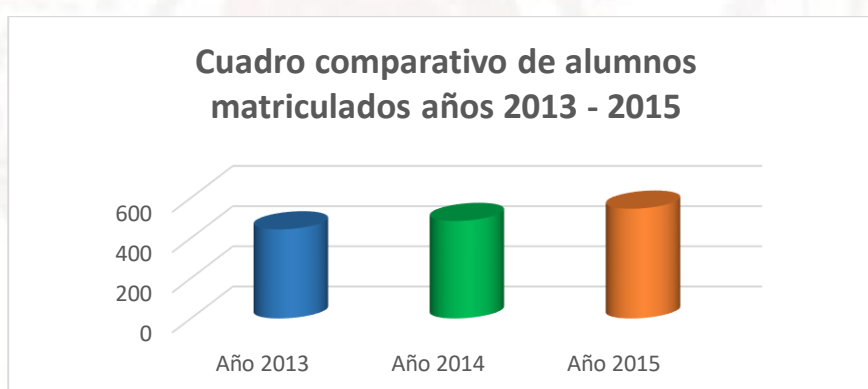
Fuente: Registros de la Institución Educativa Harvard

Elaboración: Propia

En este año 2015 la cantidad de estudiantes se ve incrementado dándonos como resultado total de 546 alumnos, y con un promedio por semestre de 46 estudiantes, en el siguiente gráfico podemos ver dicho incremento de estudiantes:

De acuerdo al gráfico anterior se observa que la cantidad de alumnos ha subido una vez más que en su totalidad es de 546 alumnos con un promedio de 46 estudiantes por semestre; en el siguiente cuadro se hará una comparación de cantidades de alumnos con la finalidad de tener un mejor panorama.

Gráfico N° 04: Comparación de datos recolectados años 2013 - 2015



	Año 2013	Año 2014	Año 2015
Alumnos	443	485	546

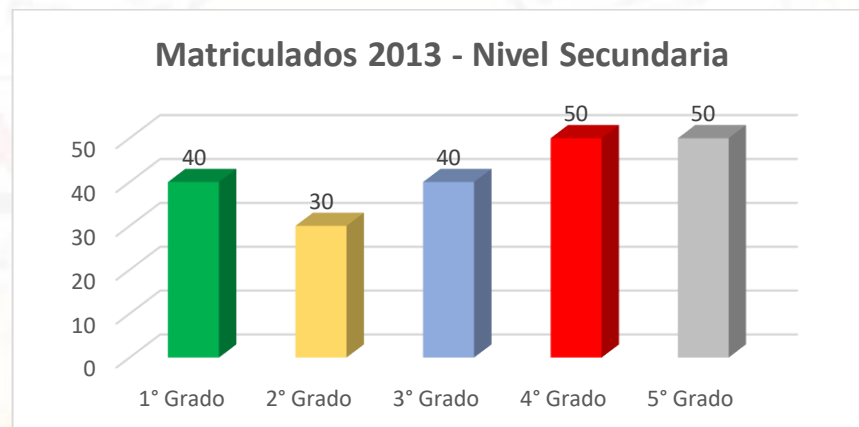
Fuente: Registros de la Institución Educativa Harvard

Elaboración: Propia

En el cuadro comparativo podemos apreciar que existe una tendencia de crecimiento en la I. E. Harvard por lo que es importante recordar que el crecimiento se da a nivel académico, pero a razón de ello los directivos muestran una preocupación por el balance que se debe de alcanzar entre el crecimiento y el desarrollo.

La Institución Educativa Harvard cuenta también con el nivel secundario, que al igual que el nivel primario se ha notado un regular incremento de alumnos durante los últimos 3 años (2013 – 2015), por lo que para una mayor información detallamos a los matriculados para el año 2013:

Gráfico N° 05: Datos recolectados correspondiente al nivel secundario año 2013



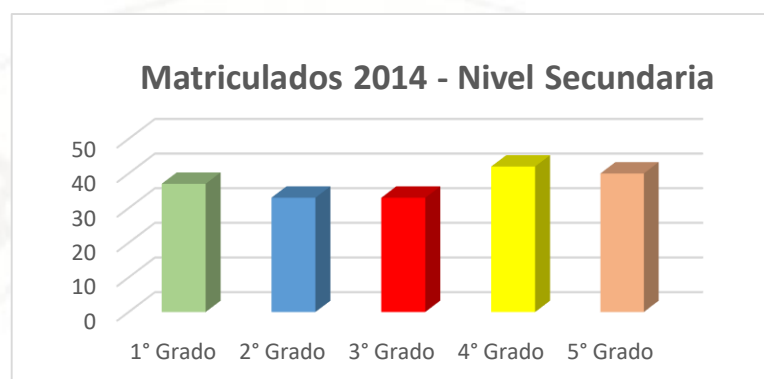
	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado
Alumnos	40	30	40	50	50

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

En este año 2013 se pudo recolectar un número considerable de alumnos matriculados que en total se llega a 210 alumnos, también se puede notar que tanto en el 4° y 5° grado del nivel secundario existe mayor concentración de alumnos.

En el siguiente cuadro se puede ver los datos con respecto al año 2014 y que a continuación se detalla:

Gráfico N° 06: Datos recolectados correspondiente al nivel secundario año 2014

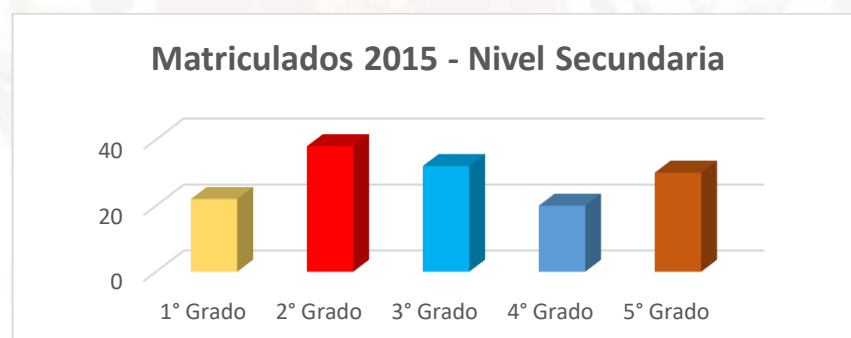


	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado
Alumnos	37	33	33	42	40

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

Como se puede apreciar en el cuadro anterior, para el año 2014 no se ha podido mantener este número de alumnos matriculados en comparación con el año 2013, muy por el contrario, se ha notado una disminución de alumnado. Véase en el gráfico, que tanto en el cuarto y quinto grado de secundaria se tuvo una baja para este año, una fuerte disminución de alumnado, a comparación con años anteriores se tuvo una baja de estudiantes llegando a tener 185 alumnos para el año 2014 y que se puede ver que tanto en el cuarto y quinto grado hubo una deserción estudiantil, y frente a ello no se pudo plantear ninguna estrategia. Para el siguiente cuadro se ha podido recolectar la siguiente información:

Gráfico N° 07: Datos recolectados correspondiente al nivel secundario año 2015



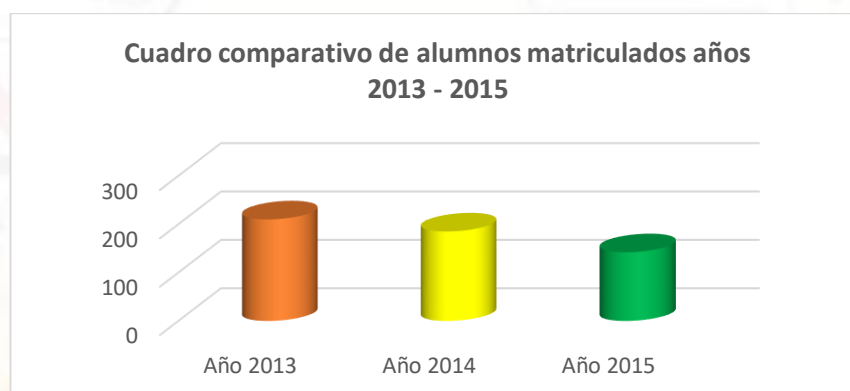
	1° Grado	2° Grado	3° Grado	4° Grado	5° Grado
Alumnos	22	38	32	20	30

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

Del cuadro anterior para el año 2015, solamente se pudo mantener algo de los estudiantes para cada grado del nivel secundario. Si comparamos entre los años 2013 con respecto a los años 2014 y 2015 se observa que existen algunos grados donde se ha mantenido el número de alumnos matriculados, pero también en algunos se ha visto cierta baja de estudiantes y muy notorio en los últimos grados del nivel secundario.

Al igual que el nivel primario, se va a describir el comportamiento anual de los alumnos matriculados en la I. E. Harvard dentro de los últimos 3 años, esto se puede observar en el siguiente cuadro:

Gráfico N° 08: Comparación de datos recolectados años 2013 - 2015



	Año 2013	Año 2014	Año 2015
Alumnos	210	185	142

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

Del gráfico anterior, se puede apreciar que para los últimos años se ha visto una baja muy llamativa en cuanto a la cantidad de alumnos matriculados; la intención de los directivos es posicionarse en el mercado educativo y a la vez incrementar la cantidad de estudiantes.

En la actualidad y hablando con los directivos, ellos se encuentran muy satisfechos con el crecimiento general (nivel primario y secundario) que el colegio ha logrado y que a través del cuadro siguiente se puede apreciar.

Gráfico N° 09: Comparación de datos recolectados años 2013 – 2015



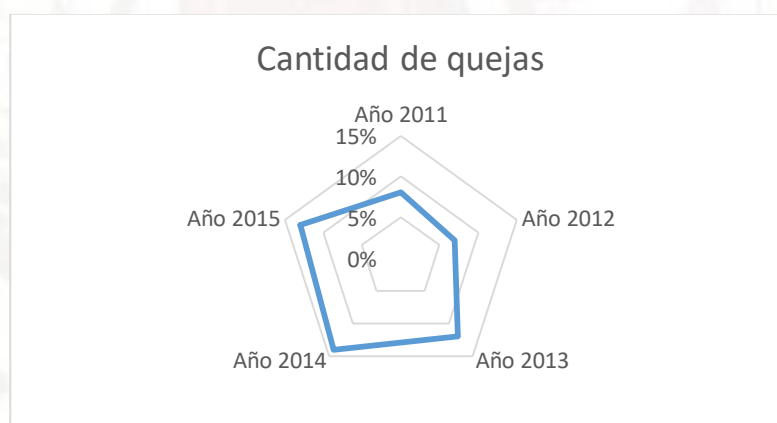
	Año 2013	Año 2014	Año 2015
Alumnos	653	670	688

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

De acuerdo al cuadro anterior se observa que el colegio ha evolucionado en cuanto a la cantidad de estudiantes, esto si lo queremos ver en ambos niveles, pero podemos apreciar que si separamos por niveles estamos viendo una baja en el nivel secundario y muy posible baja en el nivel primario a futuro.

Adicionalmente, los directivos han podido notar una serie de quejas presentadas tantas por los alumnos, así como por los padres de familia, dichas quejas también lo hemos podido recolectar y lo plasmamos en el cuadro siguiente:

Gráfico N° 10: Comparación de datos recolectados años 2013 - 2015



	Año 2011	Año 2012	Año 2013	Año 2014	Año 2015
Cantidad de quejas	8%	7%	12%	14%	13%

Fuente: Registros de la Institución Educativa Harvard
Elaboración: Propia

Para la construcción del gráfico anterior se ha tomado en consideración muchos aspectos y que se puede mencionar:

- Bajo rendimiento académico por parte de los alumnos.
- Faltas o tardanzas de los profesores.
- Faltas por parte de los alumnos.
- Quejas por maltrato por parte de los docentes hacia los alumnos.

Pese a que la institución ha implementado su libro de reclamaciones con la finalidad de lograr la eficiencia y salvaguardar los derechos de los clientes, todo esto sustentado en el Decreto Supremo N° 042-2011-PCM, artículo 3 en el que determina que las entidades de la administración pública, señaladas en los numerales de 1 a 7 del artículo I del Título Preliminar de la Ley N° 27444, deben contar con el referido libro de reclamaciones, para recibir y consignar los reclamos de los usuarios, así como para ser consultado por éstos para corroborar el tipo de prestaciones que brindan. Aun así, teniendo este libro de reclamaciones las quejas se han dado en forma verbal por parte de nuestros clientes hacia los directivos de la institución.

1.2 Formulación del Problema

1.2.1 Problema General

¿Podría un modelo de minería de datos influir en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo?

1.2.2 Problemas Específicos

- a) ¿Podría un modelo de minería de datos influir en la mejora de las **condiciones básicas de Infraestructura** en la Institución Educativa Particular Harvard-Huancayo?
- b) ¿Podría un modelo de minería de datos influir en la mejora del **Proceso enseñanza aprendizaje** en la Institución Educativa Particular Harvard-Huancayo?

1.3 Objetivos

1.3.1 Objetivo General

Determinar si un modelo de minería de datos podría influir en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo.

1.3.2 Objetivos Específicos

- a) Determinar si un modelo de minería de datos podría influir en la mejora de las **condiciones básicas de Infraestructura** en la Institución Educativa Particular Harvard-Huancayo.
- b) Determinar si un modelo de minería de datos podría influir en la mejora del Proceso **enseñanza aprendizaje** en la Institución Educativa Particular Harvard-Huancayo.

1.4 Justificación

1.4.1 Justificación Teórica:

Esta justificación es porque busca aplicar las teorías de calidad de servicio educativo para mejorar el servicio utilizando la herramienta de la minería de datos dentro de la Institución Educativa Harvard contribuyendo a la toma de decisiones de la organización.

1.4.2 Justificación Metodológica

Para el análisis y explotación de datos es necesario seguir un conjunto de pasos utilizando la metodología de minería de datos. Esta metodología busca encontrar patrones de comportamiento que ayuden a mejorar la calidad de servicio en la institución educativa Harvard, a la vez que sirve de soporte para la toma de decisiones.

1.4.3 Justificación Práctica

Como se ha podido observar, la Institución Educativa Harvard está en constantes cambios, por lo que los datos que se van almacenando también estarán en constante cambio por lo que es necesario un análisis de los datos de manera clasificada y ordenada para que así se pueda tomar algunas estrategias a futuro, el afán del presente trabajo es dar a conocer a la parte

jerárquica información relevante para la mejora de la calidad de servicio que ayudara a la toma de decisiones.



CAPITULO II

MARCO TEÓRICO

2.1 Antecedentes

En el presente capítulo se presenta algunos antecedentes de trabajos realizados que tengan relación con la situación a resolver para ello se citara cinco antecedentes en los cuales se describe como se han solucionado algunos otros problemas utilizando minería de datos, seguidamente se presenta el marco teórico en el cual se muestra todo el basamento y sostén teórico bajo el cual se enmarca el desarrollo de esta tesis, así mismo se presentará el modelo aplicativo el cual muestra una serie de pasos, etapas o fases que se tiene que seguir para la aplicación de la teoría que se haya descrito en el marco teórico.

A1. Mendoza Aguilar, Ray Israel (2009). Apoyo a la toma de decisiones con la minería de datos aplicado a una entidad del gobierno peruano. Tesis para optar por el Título de Ingeniero Informático y Sistemas. Universidad Católica del Perú. Lima. Perú. De la referida tesis se puede concluir en lo siguiente: “La minería de datos se presenta como una tecnología emergente, con varias ventajas. Por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocio; por otro ahorra grandes cantidades de dinero a una empresa y abre nuevas oportunidades negocios, logrando de esta manera que una empresa este en el rango de competentes lo cual indica estar acorde a las exigencias del mercado. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sin número de detalles debido a que el producto final involucra toma de decisiones dentro de los tres niveles es decir operativo, táctico y estratégico”.

A2. Villanueva Ojeda, Álvaro. (2008). Análisis, diseño e implementación de un datawarehouse de soporte de decisiones para un hospital del sistema de salud público. Trabajo de tesis. Pontificia Universidad Católica del Perú. Lima – Perú. Las entidades

de salud del sector público deben de tomar decisiones orientadas a satisfacer la demanda de servicios de los pacientes que acuden a los centros de salud y es por ello muy importante buscar mejorar los sistemas de información ligados a estos procesos de decisión. En este trabajo de tesis se propone la construcción de un datawarehouse que servirá de apoyo en el proceso de toma de decisiones del directorio del hospital, el cual, decidirá en base a datos históricos y cuadros generados en línea. Un sistema de este tipo permitirá reducir carga de pabellones, optimizar el uso del personal, mejorar la atención al paciente, mejorar la calidad de servicio otorgada, brindar un servicio especializado a los pacientes, gestionar recursos, conocer el estado actual de los pacientes, identificar fallas en los procesos, realizar auditorías y realizar notificaciones en tiempo real, entre otras cosas. Los datos almacenados por la entidad no tienen utilidad si es que no se transforman en información que sirva como base para tomar decisiones. Es por ello que es necesario que todos los datos históricos sean sometidos a un proceso de limpieza para poder garantizar su confiabilidad. Este sistema se encargará de hacer una limpieza de los datos almacenados para poder generar con ellos reportes que ayuden al directorio a la toma de decisiones. Para la realización del actual tema de tesis, se está optando por utilizar la suite de Inteligencia de Negocios proporcionada por Pentaho, la cual es una herramienta libre y completa. Con el uso de esta herramienta se garantiza que la entidad de salud pública no tendrá que destinar costos adicionales por licencias de software. Sin embargo, la dificultad en implementar con esta herramienta viene dada por su misma naturaleza libre (open source) y su poco tiempo en producción. Por esta razón, el presente proyecto dará pautas para la utilización e instalación de esta suite, lo cual servirá de base para proyectos similares que deseen implementar proyectos con ella.

Este aporte a la ciencia facilita la creación de un datawarehouse de manera sistemática, ordenada, e integrada dando las primeras pautas para el inicio de un proyecto basado en inteligencia de negocios.

A3. Alejandro Gutiérrez, Regina Motz, Beatriz Revello, Lydia Silva. (2010), Construcción de un sistema de apoyo a la toma de decisiones para el área de gerencia del Hospital de Clínicas. Instituto de Computación. Facultad de Ingeniería. Universidad de la República, Montevideo. Uruguay. Analizan distintos sistemas de manejo de información, resaltando las ventajas de los más avanzados frente a los primitivos, que

ofrecen mínima interacción entre el usuario y la máquina, son de respuesta lenta y no permiten respuestas a cuestiones complejas de interés para el usuario. Sistemas más modernos ofrecen mejoras, en cuanto a herramientas de análisis, mayor capacidad de manejo de información, interacciones complejas, y uso de sistemas operacionales presentes en la empresa.

Etapas 1. Adquirir conocimientos de base: datos históricos, entrevistas.

Etapas 2. Análisis de información y construcción del sistema legado (en base a sistemas operacionales ya existentes).

Etapas 3. Diseño conceptual multidimensional, en esta etapa se elige la herramienta de diseño a usar para representar el diseño conceptual del modelo de datos del datawarehouse, y el enfoque elegido para el diseño lógico del modelo de datos del datawarehouse. Se crea también una tabla de correspondencias entre el diseño conceptual del sistema legado y el diseño conceptual del datawarehouse.

Etapas 4. Diseño lógico multidimensional, en esta etapa se selecciona el enfoque a seguir en el diseño. Se diseña la arquitectura en estrella para cada relación dimensional del datawarehouse. Se crea una tabla de correspondencias entre el diseño lógico del sistema legado y el diseño lógico de la data. Se diseñan los procesos de carga y refresco.

Etapas 5. Construcción del prototipo de carga y refresco del datawarehouse.

La intención final es la de poder contar con información actualizada e inmediata para la toma de decisiones estratégicas el mencionado antecedente da pautas claras de cómo tomar decisiones bajo un sistema de apoyo, teniendo en cuenta ciertas restricciones.

2.2 Bases Teóricas

2.2.1 Inteligencia de Negocio

La Inteligencia de Negocios o Business Intelligence (BI) es la solución a ese problema, pues por medio de dicha información puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva. La clave para la inteligencia de negocios es la información y uno de sus mayores beneficios es la posibilidad de utilizarla en la toma de decisiones. En la actualidad hay una gran variedad de software de

BI con aplicaciones similares que pueden ser utilizados en las diferentes áreas de la empresa, tales como, ventas, marketing, finanzas, etc. Son muchas las empresas que se han beneficiado por la implementación de un sistema de BI, además se pronostica que con el tiempo se convertirá en una necesidad de toda empresa. En este nuevo mundo, la información reina afirma Geoffrey A. Moore, Director de Chasm Group. Vivimos en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y tomadores de decisiones requieren de un acceso rápido y fácil a información útil y valiosa de la empresa. Una forma de solucionar este problema es por medio del uso de Business Intelligence o Inteligencia de Negocios. La Inteligencia de Negocios o Business Intelligence (BI) se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos. Dentro de la categoría de bienes se incluyen las bases de datos de clientes, información de la cadena de suministro, ventas personales y cualquier actividad de marketing o fuente de información relevante para la empresa. La inteligencia de negocios apoya a los tomadores de decisiones con la información correcta, en el momento y lugar correcto, lo que les permite tomar mejores decisiones de negocios. La información adecuada en el lugar y momento adecuado incrementa efectividad de cualquier empresa. La tecnología de BI no es nueva, ha estado presente de varias formas por lo menos en los últimos 20 años, comenzando por generadores de reportes y sistemas de información ejecutiva en los 80's Afirma Candice Goodwin. Entiéndase como sinónimos de tecnología de BI los términos aplicaciones, soluciones de software de inteligencia de negocios. 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. A continuación, mostramos el gráfico N° 01 La cual ayuda a entender la jerarquía de conocimiento.

Gráfico N° 11 Jerarquía del conocimiento



Fuente: Bitam. "Business Intelligence". En línea para descargas. Disponible en: <http://www.bitam.com/spanish/AcercaDeBI.htm>. Consulta: 18 marzo del 2011.

Nótese en el gráfico anterior que el conocimiento de alto valor es el cual se encuentra dentro de las organizaciones pero en bajo volumen es por ello que intentamos incrementar este conocimiento desde un punto de vista pragmático, y asociándolo directamente con las tecnologías de información, podemos definir a través de Business Intelligence el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP/OLAP, minería de datos, datawarehouse, etc.) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones de la organización.

La Inteligencia de Negocios 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 costos, 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 Integral (CMI), Sistemas de Soporte a la Decisión (DSS), Sistemas de Información Ejecutiva.

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un datawarehouse, los datos están desnormalizados para apoyar a consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado. En este sentido los procesos de ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén totalmente integrados. En definitiva, una solución BI completa permite: Observar, Comprender, Predecir, Colaborar, Decidir.

2.2.2 Educación

La educación se define como un proceso conscientemente organizado, dirigido y sistematizado sobre la base de una concepción pedagógica determinada, que se plantea como objetivo más general la formación multilateral y armónica del educando para que se integre a la sociedad en que vive y contribuya a su desarrollo y perfeccionamiento. El núcleo esencial de esa formación ha de ser la riqueza moral. La educación es la interacción humana en la cual, sobre la base de un potencial biológico, seres humanos con la mediación de otros, se apropian la cultura que los constituye en la clase peculiar de ser humano que son, tanto en el sentido colectivo como en el personal, para crear su propia versión particular de la misma, que es su personalidad. Esa apropiación-mediatización genera las competencias humanas (significados, destrezas, actitudes) que se requieren en esa particular colectividad para ser miembro de ella y reproducirla en su estructura, fines y buen funcionamiento y transformarla a tono con los cambios que se producen en el entorno natural y cultural. La existencia y buen funcionamiento de la sociedad (estado de salud) está en directa relación con la educación. Decimos que la educación es de "calidad" cuando persigue ciertos fines sociales que estimamos valiosos y los logra de manera creativa, eficiente y efectiva. La calidad de la educación es relativa a los valores y fines sociales a los que sirve. Es así, que la educación debe ser

considera como un elemento que sirve de palanca del cambio del país, y como el medio principal para la generación de empleos, además de una participación más equitativa de la economía, del federalismo y apoyo al desarrollo regional.

2.2.3 Calidad

En la última década la calidad se ha convertido en un concepto citado por las principales instituciones públicas y de servicios, se ha convertido en una meta que es buscada de una manera completa, ya que se ha considerado de forma común que lo que tiene “calidad” cubre con las expectativas del cliente, la calidad en general abarca todas las cualidades con las que cuenta un producto o un servicio, cuando sus características, tangibles e intangibles satisfacen las necesidades del usuario, estos conceptos fueron utilizados primeramente en el rubro de la economía y área industrial, hoy en día la competitividad se presenta cada vez más en las empresas, y de esta manera estar al nivel de los estándares internacionales de calidad, este concepto es citado cada vez por las instituciones dedicadas a la educación. Dentro de los postulados de ANUIES, la calidad es considerada de una forma dinámica, lo cual determina que debe estar en continuo cambio, pero fundamentados en acciones tendientes a tener elementos suficientes para determinar niveles o cambios provocados en nuestros procesos educativos con los miembros que desarrollan dichas funciones, estudiantes, maestros y directivos.

La aplicación del concepto de calidad, relacionado al desarrollo de programas de desarrollo institucional bajo una planeación a largo plazo permite que el beneficio de estos planes toque a cada uno de los elementos que forman a la institución educativa.

Entendemos esta acepción para aplicar el grado de valor, bueno o malo, que tienen o se atribuyen a las cosas; también aceptamos esta otra con más amplitud “calidad es el conjunto de cualidades de un ser, objeto o proceso que le confieren la aptitud de satisfacer necesidades explícitas o implícitas”. Para nuestro interés, calidad en términos de servicios es más difícil su percepción que se trata de relaciones, comunicaciones y procesos y éstas forman parte

de la calidad de las transformaciones sociales, por lo tanto, vinculadas a valores, perjuicios, estereotipos, actitudes, motivaciones y comportamientos.

Calidad total educativa en el Perú

Uno de los principales desafíos para las políticas sociales en el Perú es el mejoramiento de la calidad de la educación, concepto sobre el cual no existe una definición que goce de consenso. Por lo tanto, el concepto calidad de la educación es subjetivo, socialmente construido, dependiente del contexto y está integrado por valores. De esta manera, distintos actores de la sociedad, tales como docentes, padres de familia, estudiantes y empresarios, manejan conceptos distintos.

La calidad educativa es pues un proceso que nos debe dirigir hacia la idealidad. No tratamos de soñar sino de materializar aquello que tenemos como idea.

Entonces para lograr esa calidad educativa deben intervenir entre otros: El Estado, la Municipalidad, los profesores, los alumnos, la comunidad, los padres de familia, los medios de comunicación social (prensa, radio, tv.) y todas las instituciones que rodean el entorno social.

Ahora bien, en el mejoramiento de la calidad de la educación, se tiene necesariamente considerar que en este trabajo se cuenta con los sujetos (educandos, educadores y comunidad), elementos (objetivos, contenidos, métodos, medios, materiales, infraestructura, tiempo) y procesos (planeamiento, investigación, programación, implementación, ejecución y evaluación) de calidad, a este conjunto podemos llamarle calidad total.

Por ello, se puede entender calidad educativa como el proceso consensuado (por todos los miembros implicados) de construcción de objetivos para cada contexto y momento y como la trayectoria o tendencia que realizamos para conseguirlo. Es un proceso de construcción y/o fijación (por miembros de los distintos estamentos de la comunidad educativa) de objetivos educativos, no sólo referidos a los logros de los alumnos, sino también a cuestiones curriculares y organizativas, a la vida del centro y el camino que recorreremos para lograrlos. Ahora bien, existen algunas consideraciones para una calidad

educativa: Si uno o más intervinientes en este proceso no reúnen condiciones de calidad, aun así, los demás sean excelentes, de seguro que el producto final será deficiente. En cambio, si cada uno de ellos (sujetos, elementos y procesos) es de calidad se hace innecesaria la supervisión o control y los productos finales son cada vez más óptimos. La calidad debe concretarse en el producto final es decir el alumno, pero como responsabilidad de los agentes educativos, beneficia a la institución y a la comunidad en su conjunto. Finalmente, el elemento más valioso de la humanidad ya no es el dinero, ni los recursos naturales, sino la educación.

2.2.4 Pilares de la Calidad Total

Uno de los pilares de la calidad total es la organización ya que la organización es el aula, el equipo docente-alumno es equivalente a los trabajadores. El alumno es el cliente del docente que recibe los servicios educativos para su desarrollo. La institución y los docentes son los proveedores. Cada uno en la organización, debe dedicarse al mejoramiento común, continuo, personal y colectivamente. Debe buscarse un ambiente de aprendizaje en colaboración y asignarse recursos, tiempo, dinero y potenciales para su capacitación permanente. Los alumnos no deben concentrarse en las notas sino en el aprendizaje. La organización es un sistema y el trabajo que las personas realizan es como un proceso continuo. Cada sistema está conformado por procesos, las mejoras en la calidad de estos procesos determinan la calidad en los productos resultantes mientras que el mejoramiento continuo de los procesos de aprendizaje reemplaza a "enseñar y evaluar".

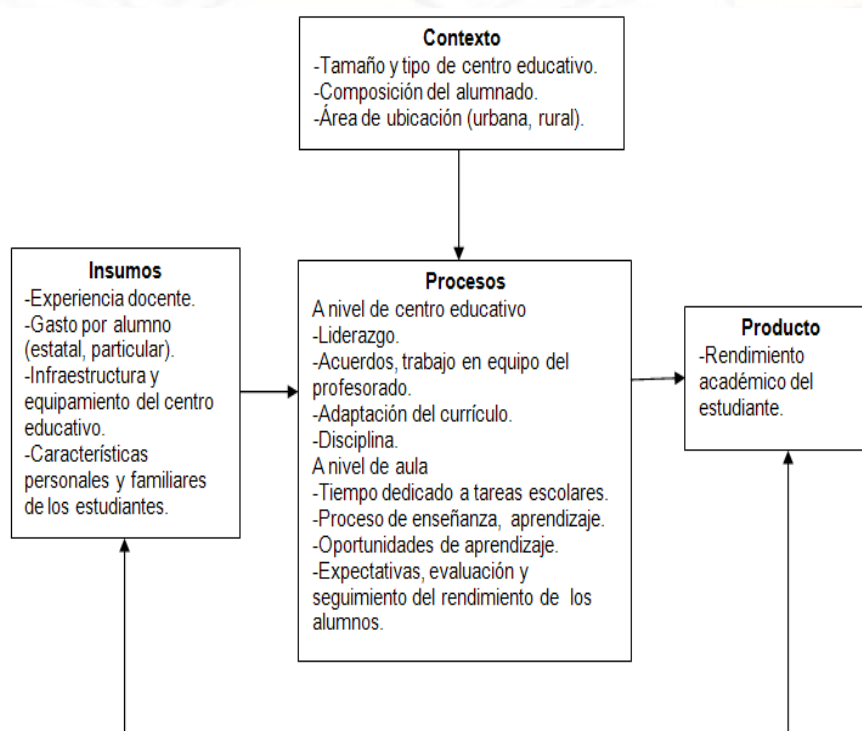
El éxito de la calidad es responsabilidad de la gerencia este debe preocuparse que la calidad total se internalice en la cultura de la organización y en crear constancia de propósitos para el mejoramiento del producto y del servicio. Alumnos y docentes alcanzan su máximo potencial a través del mejoramiento continuo de sus labores.

El enfoque de insumos, procesos y resultados educativos

En el Perú, al igual que en algunos otros países, se ha propuesto una visión de la calidad educativa basada en un análisis de los insumos disponibles, los

procesos educativos y los resultados obtenidos en diversos contextos. Se entiende por insumos a los recursos disponibles que tienen impacto directo en el nivel de logro o rendimiento académico del alumno, al cual se le denomina producto. Los procesos son aquellos elementos que median en la relación entre insumos y productos, mientras que el contexto se refiere al ambiente en el que se desenvuelve la institución. En el gráfico 02 se presentan algunos ejemplos de los factores que pueden considerarse en cada una de estas dimensiones.

Gráfico N° 12 Modelo de calidad educativa



Fuente: Gestión educativa en el Perú.

<http://www.monografias.com>

Consulta: 18 marzo del 2012.

Desde un enfoque económico, puede decirse que el nivel de logro de los estudiantes o producto es el resultado de la interacción de la oferta y demanda educativa. En donde se entiende por oferta a las características de los docentes y directores, la infraestructura del centro educativo, entre otros; y por demanda, a lo relacionado con las características socioeconómicas de los alumnos y su familia. Las investigaciones recientes que tratan de explicar el

nivel de logro o rendimiento académico de los alumnos, a partir de los insumos, procesos y el contexto, han enfatizado la identificación de aquellos aspectos que favorecen la reducción de las brechas en el rendimiento académico que se originan por desigualdades socioeconómicas. Este interés surge como respuesta a algunas investigaciones que sugerían que, en las escuelas, básicamente, se reproducían las desigualdades de origen social: donde los alumnos de un bajo nivel socioeconómico y cultural obtenían los peores resultados educativos.

Modelo de la estimación de la calidad en la educación

En los últimos años han surgido diversos modelos de estimación de la calidad aplicables a cualquier tipo de empresa u organización, tanto del sector privado como del público, hospitales, centros de educación, pequeñas organizaciones, etc. los mismos que no resultan incompatibles entre sí, sino que buscan poner en marcha procesos de articulación entre todos los componentes de la organización que permitan la gestión y la conducción de la calidad. Estos modelos no se limitan a llevar a cabo mediciones, se trata más bien, de procesos dinámicos que buscan fundamentalmente analizar e interpretar la información a partir de un marco referencial que permita emitir juicios de valor y tomar decisiones.

La ISO es una herramienta para la gestión de la calidad que promueve un enfoque basado en procesos estandarizados para aumentar la satisfacción del cliente mediante el cumplimiento de sus requisitos.

Es un sistema que determina las necesidades y expectativas del producto o servicio. Su enfoque es basado en procesos. Realiza permanentemente una evaluación, autoevaluación y auditorías de eficacia. Promueve la mejora continua y utiliza las técnicas estadísticas. Además, dicho modelo permite configurar un sistema de gestión de la calidad. Los sistemas de gestión de calidad tienen dos características: Pueden mejorar continuamente y Evolucionan. El gráfico N° 2.3 muestra el modelo del sistema de gestión de calidad basado en el ISO 9001.

Gráfico N° 13 Modelo del sistema de gestión de calidad basado en el ISO 9001



Fuente: Gestión educativa en el Perú.

<http://www.monografias.com>

Consulta: 18 marzo del 2012.

Nótese que el gráfico anterior está basado en el ISO 9001, para lo cual es importante recordar que una ISO es una red de los institutos de estandarización nacionales de 157 países, sobre la base de un miembro por país, con una secretaría Central en Ginebra, Suiza, que coordina el sistema. Es una organización no gubernamental: sus miembros no son del sistema de Naciones Unidas, o delegaciones de gobiernos nacionales. Sin embargo, la ISO tiene un puesto especial entre el sector público y sectores privados.

Globalización: Calidad Educativa y Profesionales Competitivos

La competitividad de una nación depende de la capacidad de las personas para innovar y perfeccionarse. La globalización económica ofrece la oportunidad de integrar la fuerza laboral de los países en vías de desarrollo, pero esto implica el reto de vigilar la calidad educativa de la formación en los diferentes niveles educativos y propuestas de profunda reestructuración en búsqueda de otorgar a nuestros jóvenes las competencias tecnológicas, humanísticas y emocionales necesarias para hacerlos competitivos en el mercado global.

En nuestro país hubo un mal enfoque del sistema educativo respecto a las necesidades de la sociedad. Se creó una ruptura entre la formación y las

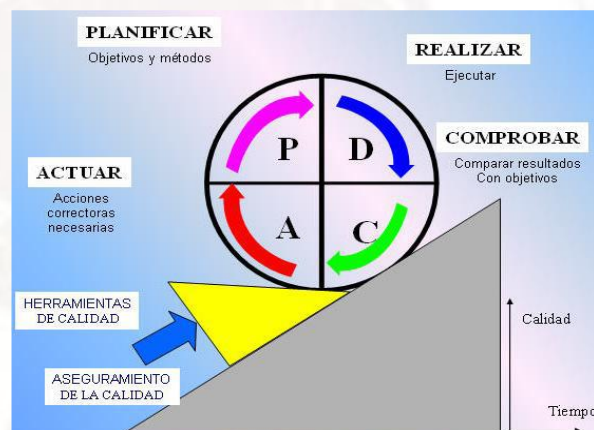
exigencias del mercado laboral, por un lado, y por otro la poca investigación realizada y el sistema productivo.

Un análisis del perfil de nuestros profesionales muestra las siguientes carencias: Son memoristas, no razonan lógicamente, no leen, no saben redactar un ensayo, no tienen cultura general, tienen miedo al fracaso, sin el profesor no saben qué hacer, tienen inculcada la mentalidad del empleado que hace lo que el jefe le indica.

En respuesta a esto las Instituciones de Educación Superior (IES) vienen registrando importantes avances en el desarrollo de sus sistemas de gestión de la calidad, para lo cual aplican algunos modelos que poseen en la actualidad alto reconocimiento internacional como son: Normas ISO, especialmente las de la familia ISO 9000, sistemas de Acreditación de Programas e Instituciones que se han desarrollado particularmente en el ámbito de la educación superior, premios nacionales e internacionales de calidad, que en algunos casos incluyen Galardones de Excelencia para la gestión escolar.

Por otro lado, a diferencia de la forma convencional de gestionar la educación a través de un ciclo abierto, la búsqueda de la calidad en la educación lleva a la aplicación del ciclo Deming (Plan=Planear, D=Hacer, C=Verificar, A=Actuar) como parte de un plan sostenido de mejora continua, utilizando herramientas de medición donde participan los profesores, alumnos, egresados y representantes de las empresas.

Gráfico N° 14 muestra el Ciclo de mejora continua de Deming.



Fuente: Ciclo de mejora continua de Deming.

<http://cmzconsulting.blogspot.com>

Consulta: 18 marzo del 2012.

Del gráfico anterior es importante saber las reglas básicas para la mejora continua son que: no se puede mejorar nada que no se haya controlado, no se puede controlar nada que no se haya medido, no se puede medir nada que no se haya definido y por ultimo no se puede definir nada que no se haya identificado.

2.2.5 Fundamentos de la Minería de Datos

Las técnicas de minería de datos son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. La minería de datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. La minería de datos está lista para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras: Recolección masiva de datos, potentes computadoras con multiprocesadores, algoritmos de minería de datos. Por otro lado, es importante rescatar las principales características y objetivos de la minería de datos las cuales se describen a continuación:

- El entorno de la minería de datos suele tener una arquitectura cliente servidor.
- Las herramientas de la minería de datos ayudan a extraer el mineral de la información enterrado en archivos corporativos o en registros públicos, archivados.
- El minero es, muchas veces un usuario final con poca o ninguna habilidad de programación, facultado por exploradores de datos y otras poderosas herramientas indagatorias para efectuar preguntas y obtener rápidamente respuestas.
- Hurgar y sacudir a menudo implica el descubrimiento de resultados valiosos e inesperados en un cúmulo de datos dispersos los cuales se tienen que manipular de tal manera que se obtenga a partir de datos

información y posteriormente a partir de esa información obtener conocimiento la cual será útil para el proceso de toma de decisiones.

- Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente debido a los potentes algoritmos con lo que se cuenta.
- Debido a la gran cantidad de datos, algunas veces resulta necesario usar procesamiento en paralelo para la minería de datos, esto se puede lograr siguiendo el proceso de minería de datos el cual sigue un ciclo flexible que permite retornar a pasos anteriores en caso de que exista la necesidad.
- La minería de datos produce cinco tipos de información: asociaciones, secuencias, clasificaciones, agrupamientos, pronósticos.
- Los usuarios finales (mineros de datos) utilizan varias herramientas y técnicas las cuales les brinda una interfaz amigable y comprensible de tal manera que no exista la necesidad de que dichos usuarios sean expertos en programación u otros aspectos, ya que solo se limitaran a usar las bondades que la minería de datos brinda.

La minería de datos es un proceso que invierte la dinámica del método científico en el siguiente sentido: En el método científico, primero se formula la hipótesis y luego se diseña el experimento para coleccionar los datos que confirmen o refuten la hipótesis. Si esto se hace con la formalidad adecuada (cuidando cuáles son las variables controladas y cuáles experimentales), se obtiene un nuevo conocimiento. En la minería de datos, se coleccionan los datos y se espera que de ellos emerjan hipótesis. Se busca que los datos describan o indiquen por qué son como son. Luego entonces, se valida esa hipótesis inspirada por los datos en los datos mismos, será numéricamente significativa, pero experimentalmente inválida. De ahí que la minería de datos debe presentar un enfoque exploratorio, y no confirmador. Usar la minería de datos para confirmar las hipótesis formuladas puede ser peligroso, pues se está haciendo una inferencia poco válida. La minería de datos es una tecnología compuesta por etapas que integra varias áreas y que no se debe

confundir con un gran software. Durante el desarrollo de un proyecto de este tipo se usan diferentes aplicaciones software en cada etapa que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente. Actualmente existen aplicaciones o herramientas comerciales de minería de datos muy poderosas que contienen un sinfín de utilerías que facilitan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta.

a) Alcances de la minería de datos

El nombre de minería de datos deriva de las similitudes entre buscar valiosa información de negocios en grandes bases de datos, por ejemplo, encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de minería de datos puede generar nuevas oportunidades de negocios al proveer capacidades como son: Predicción automatizada de tendencias y comportamientos. La minería de datos automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerirían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (targeted marketing). La minería de datos usa datos en mailing promocionales anteriores para identificar posibles objetivos para maximizar los resultados de la inversión en futuros mailing. Otros problemas predecibles incluyen pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que probablemente respondan similarmente a eventos dados.

El descubrimiento automatizado de modelos previamente desconocidos es una de las ventajas que nos ofrece la minería de datos ya que las

herramientas de minería de datos barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otro problema de descubrimiento de modelos incluye detectar transacciones fraudulentas de Tarjetas de créditos e identificar datos anormales que pueden representar errores de tipado en la carga de datos.

Las técnicas de minería de datos pueden redituar los beneficios de automatización en las plataformas de hardware y software existentes y puede ser implementada en sistemas nuevos a medida que las plataformas existentes se actualicen y nuevos productos sean desarrollados. Cuando las herramientas de minería de datos son implementadas en sistemas de procesamiento paralelo de alta performance, pueden analizar bases de datos masivas en minutos. Procesamiento más rápido significa que los usuarios pueden automáticamente experimentar con más modelos para entender datos complejos. Alta velocidad hace que sea práctico para los usuarios analizar inmensas cantidades de datos y grandes bases de datos, a su vez, producen mejores predicciones.

b) Toma de decisiones

Una decisión es una elección consciente y racional, orientada a conseguir un objetivo, que se realiza entre diversas posibilidades de actuación (o alternativas). Antes de tomar una decisión deberemos calcular cual será el resultado de escoger una alternativa. En función de las consecuencias previsibles para cada alternativa se tomará la decisión. Así, los elementos que constituyen la estructura de la decisión son: los objetivos de quién decide y las restricciones para conseguirlos; las alternativas posibles y potenciales; las consecuencias de cada alternativa; el escenario en el que se toma la decisión y las preferencias de quien decide. En la toma de decisiones se destacan algunas etapas dichas etapas están ilustradas en el gráfico N° 05.

Gráfico N° 15 Etapas de la toma de decisiones



Tomado de: HAROLD KOONTZ y HEINZ WEHRICH "Administración una Perspectiva Global" McGRAW- HILL INTERAMERICANA DE EDITORES, S.A., 11va edición, 1999.

Por otra parte, cabe resaltar que existen dos tipos de decisiones que se detallan a continuación:

- **Decisiones Programadas**

Una decisión programada se aplica a problemas estructurados o de rutina. Los operadores de tomos por ejemplo tienen especificaciones y reglas que les señalan si la pieza que han hecho es aceptable, si tiene que desecharse o si se tienen que procesar de nuevo.

- **Decisiones No Programadas**

Las decisiones no programadas se usan para situaciones no programadas, nuevas y mal definidas, de naturaleza no repetitiva. Tenemos que tener en cuenta que la mayoría de las decisiones no son ni completamente programadas ni completamente no programadas; son una combinación de ambas. La mayor parte de las decisiones no programadas las toman los gerentes del nivel más alto esto es porque los gerentes de este nivel tienen que hacer frente a los problemas no estructurados.

c) Información e inteligencia de negocios en organizaciones

En los últimos años, ha existido un gran crecimiento en las capacidades de generar y coleccionar datos, debido básicamente al gran poder de

procesamiento de las máquinas como a su bajo costo de almacenamiento. Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información oculta, de gran importancia estratégica, a la que no se puede acceder por las técnicas clásicas de recuperación de la información. El descubrimiento de esta información oculta es posible gracias a sofisticadas técnicas donde se aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad, pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) que se encarga de la preparación de los datos y la interpretación de los resultados sostenidos, los cuales dan un significado a estos patrones encontrados.

Así el valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejorar nuestra comprensión de los fenómenos que nos rodean. Hoy, más que nunca, los métodos analíticos avanzados son el arma secreta de muchos negocios exitosos. Empleando métodos analíticos avanzados para la explotación de datos, los negocios incrementan sus ganancias, maximizan la eficiencia operativa, reducen costos y mejoran la satisfacción del cliente. Por otro lado, desde que se inició la era de la computadora, las organizaciones han usado los datos desde sus sistemas operacionales para atender sus necesidades de información. Algunas proporcionan acceso directo a la información contenida dentro de las aplicaciones operacionales. Otras, han extraído los datos desde sus bases de datos operacionales para combinarlos de varias formas no estructuradas, en su intento por atender a los usuarios en sus necesidades de información.

Ambos métodos han evolucionado a través del tiempo y ahora las organizaciones manejan una data no limpia e inconsistente, sobre las cuales, en la mayoría de las veces, se toman decisiones importantes.

El fin del siglo XX y el comienzo del siglo XXI encuentra a las empresas frente a un escenario en que la competencia es fenomenal y en el que los negocios necesitan apoyarse indefectiblemente en información útil que

los conduzcan a una mejor toma de decisiones para de esta manera lograr un posicionamiento en el mercado, satisfacción de clientela, mayores ganancias entre otros. Estos objetivos potencian la importancia de la información como recurso básico para el desarrollo competitivo de las empresas. Una de las preocupaciones principales de los directivos de esta época es cómo lograr la coherencia y compatibilidad entre las metas corporativas y las de los sistemas de información. Entre éstas, se destaca lograr la mayor utilidad de la información disponible. Obsérvese que hemos mencionado “disponible” no como el hecho de tener millones de datos almacenados sino más bien de tener información disponible de dichos datos.

Por otro lado, las empresas modernas quieren mantenerse a la vanguardia, sabemos la importancia fundamental de tener el derecho a la información, a su disposición para ayudar a tomar decisiones clave que afectan a la dirección de sus movimientos comerciales. Así, en una encuesta realizada el 2006 por Gartner Group, que es la empresa líder en investigación de tecnología y asesoramiento, encontró que el CIO y el Business Intelligence (BI) son las dos mayores tecnologías a usarse por los ejecutivos. Este énfasis se hizo eco de una encuesta realizada el 2007 por Information Week, es una revista semanal en línea sobre instigaciones de tecnología, donde se encontró que cerca de la mitad de los ejecutivos planeó aumentar su inversión en Business Intelligence (BI) por encima de los niveles del 2006; en una investigación realizada por Óscar Alonso y Pilar Pedrosa, Penteo, en el cual se muestra que en el año 2012 más del 40% de las compañías entrevistadas tienen previsto incrementar el gasto destinado a iniciativas de Business Intelligence & Performance Management. Las empresas reconocen la aportación de valor al negocio de estas herramientas, pero para ello es importante cuidar tanto el proceso de selección como el de implantación de la solución; el gráfico N° 06 se muestra los principales beneficios que las compañías consiguen con el uso de Business Intelligence (BI).

Gráfico N° 16 ¿Cuáles son los principales beneficios que su compañía consiguió con la inversión en BI?



Disponible en: <http://www.computing.es/Tendencias/201102230029/GESTION-Business-Intelligence---Performance-Management-2010.aspx>

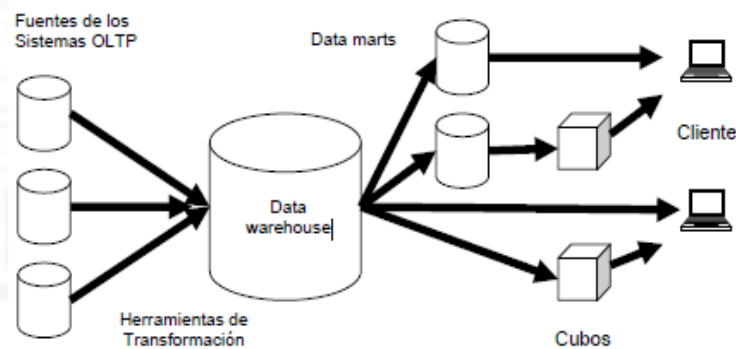
Estas tendencias indican claramente que las empresas reconocen que su capacidad de competir en gran medida depende del uso inteligente de la tecnología y la información, si el objetivo es ayudar a sus clientes a comunicarse, socializar, compartir y buscar información, entretenerse o comprar bienes y servicios de una manera más eficiente.

d) DataWarehouse

Un DataWarehouse hace referencia a un almacén de información temática orientado a cubrir las necesidades de aplicaciones de los sistemas de Soporte de Decisiones (DSS) y de la Información para Directivos (EIS), que permite acceder a la información corporativa para la gestión, control y apoyo a la toma de decisiones. Los datos en un datawarehouse tienen un ciclo de vida. Estos son originados en otros sistemas, usualmente en sistemas OLTP. Luego son transformados e importados al datawarehouse.

Una vez que estos se encuentran en el Datawarehouse, es posible construir cubos, los cuales son vistas multidimensionales recalculado de datos. Los cubos proveen velocidad y facilidad de acceso a las herramientas de consulta y por lo tanto a los usuarios finales. El gráfico N° 07. Muestra las partes del ciclo de vida cuando el datawarehouse está comenzando a crearse.

Gráfico N° 17 Ciclo de vida del Datawarehouse



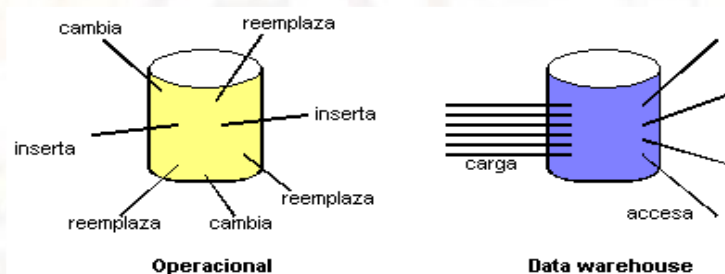
Fuente: Microsoft Corporation. Microsoft Sql Server. Datawarehousing Training.
Disponible en: www.microsoftsqlserver.com. Consulta 20 de marzo del 2011.

Nótese que toda la información dispersa de las fuentes de sistemas OLTP se intentan integrar en un solo almacén de datos gracias a este ciclo de vida un datawarehouse logra tener las siguientes características que se describen a continuación:

- **Orientado a temas:** Un DataWarehouse clasifica la información en base a los aspectos que son de interés para la empresa, de esta manera los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones.
- **Integración:** El aspecto más importante del ambiente DataWarehousing es que la información encontrada al interior está siempre integrada. La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

- **Variante:** Toda la información del DataWarehouse es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted accede a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso como la información en el DataWarehouse es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".
- **No Volátil:** La información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable. Esto se puede observar en el gráfico N° 08.

Gráfico N° 18 Volatilidad de información



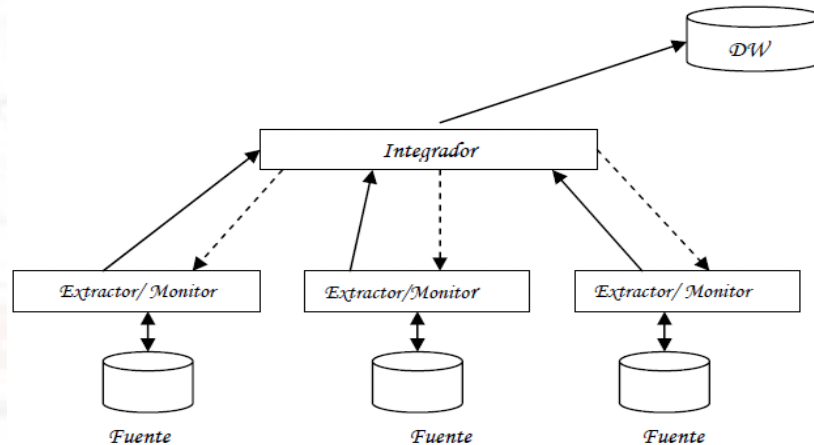
Fuente: Claudio Cesares. Data Warehousing. Disponible en <http://personal.lobocom.es/claudio/gen006.htm>. consulta: 22 marzo del 2011.

Se puede notar que normalmente en el ambiente operaciones la data es actualizada registro por registro mientras que en un DataWarehouse la data es cargada en el depósito de datos y es accesada, pero una vez que el snapshot está hecho, los datos en el depósito no cambian esto hace que no la información no sea volátil.

Por otro lado, es importante resaltar la arquitectura del datawarehouse que es la descripción de sus componentes y servicios, además de las relaciones entre ellos y su desarrollo, estas descripciones deben contener suficiente

información para permitir implementar un datawarehouse. Así mismo, el DataWarehouse tiene una arquitectura que puede ser visto como un proceso de cuatro pasos los cuales lo podemos apreciar en el gráfico N° 09.

Gráfico N° 19 Estructura básica de un DataWarehouse.



Basado en: Ralph Kimball. 1996. "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses". 3era edición. John Wiley & Sons. pag 59. ISBN 0-471-15337-0.

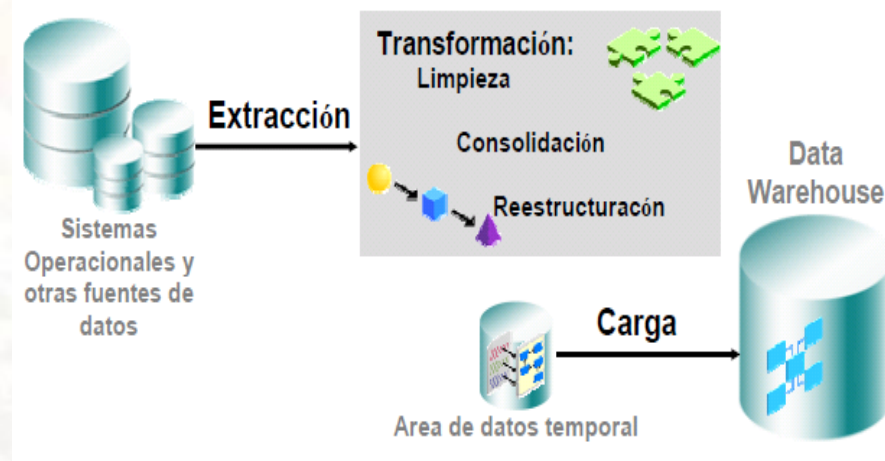
El diseño, se refiere al proceso de elegir y diseñar la estructura interna, la misma que soportara la manipulación y recuperación de los datos, es decir, diseñar la estructura interna del DataWarehouse y cómo va almacenar los datos a la par de cómo los hará disponibles.

La construcción, se refiere al diseño e implementación de las herramientas encargadas de llevar los datos de las fuentes al repositorio, y ya una vez dentro de este los homogenizará. También tenemos el Análisis que es la explotación que se realiza de los datos almacenados dentro del DataWarehouse a través de técnicas que facilitan y hacen más eficiente la consulta de los mismos. Ya con el DataWarehouse poblado lo último es diseñar e implementar una interfaz que le permita al usuario final interactuar con el sistema, brindándole todas las ventajas del análisis de la información. Y por último tenemos el mantenimiento proceso encargado de llevar los datos nuevos que se ingresen en el repositorio y también se encarga de actualizar los datos previamente calculados en las vistas.

e) Proceso Extracción, Transformación y Carga de datos

El proceso ETL proviene de las siglas extracción, transformación y carga de datos como sabemos los datos recopilados en las diferentes fuentes de almacenamiento son considerados datos “sucios” por contener muchas de las veces información innecesaria para el análisis dentro del DataWarehouse y por tener desigualdad de criterio para describir cierto dato. La limpieza de los datos “sucios” es un proceso multifacético y complejo que consiste en analizar los datos corporativos para descubrir inexactitudes, anomalías y transformar los datos para asegurar que sean precisos y coherentes. Los Sistemas ETL realizan las funciones de extracción de las fuentes de datos (transaccionales o externas), transformación (limpieza, consolidación, etc.) y la carga de los datos una vez que ya hayan pasado por los dos anteriores pasos, para así almacenarse en el DataWarehouse. El gráfico N° 10. Nos ilustra de manera gráfica el proceso ETL.

Gráfico N° 20 Proceso ELT (Extracción, Transformación y Carga)



Tomado de: LARDENT, Alberto, R. Sistemas de información para la gestión empresarial: planeamiento, tecnología y calidad. 1era. Edición. Editorial. Pearson. Buenos Aires. 2001. Capítulo 18. 544p.

En el gráfico anterior se muestra la secuencia que el proceso ETL ejecuta para realizar el mantenimiento del DW, en el cual el primer proceso es la extracción de los datos necesarios para el análisis de la información, una

vez hecho la extracción el proceso siguiente es la transformación de los datos en datos basados según el formato estandarizado dentro del DW y por último se realiza la carga de los datos en el DW. Los pasos para el proceso ETL consisten en: Validar los datos que usa la aplicación del DW para realizar las consultas de prueba. Producir la metadata, una descripción del tipo de datos, formato y el significado relacionado al negocio de cada campo.

La documentación del proceso completo, para que se pueda ampliar, modificar y arreglar los datos en el futuro con más facilidad. La carga de los datos en el DataWarehouse.

2.3 Hipótesis

2.3.1 Hipótesis General

Un modelo de minería de datos si influye en la mejora del **servicio académico** en la Institución Educativa Particular Harvard-Huancayo.

2.3.2 Hipótesis Específicas

- a) Un modelo de minería de datos **si influye** en la mejora de **las condiciones básicas de Infraestructura** en la Institución Educativa Particular Harvard-Huancayo.
- b) Un modelo de minería de datos si influye en la mejora del **Proceso enseñanza aprendizaje** en la Institución Educativa Particular Harvard-Huancayo

2.4 Definición de Términos

- a) **Business Intelligence**: Describe la habilidad de la empresa para acceder y explorar información, para desarrollar perspicacia entendimiento y mejorar la toma de decisiones.
- b) **Cubo OLAP**: Una estructura matricial de información con múltiples entradas que permiten realizar distintas combinaciones de sus elementos para visualizar resultados desde distintas perspectivas y variando los niveles de detalle.

- c) **DataMarts:** Es un subconjunto de contenidos del Data Warehouse y posee datos enfocados a nivel departamental o de un área de negocio específica.
- d) **DataWarehouse (DW):** Es un almacén de datos categorizados, que concentra información relevante, la cual es distribuida por diversas herramientas de consultas y creación de informes orientados a la toma de decisiones.
- e) **DataWarehousing:** Proceso asociado con la recolección, administración y uso de los activos de información de una empresa. Es la metodología para el diseño y administración del DataWarehouse.
- f) **Diagrama Copo de Nieve: (SnowFlake):** Es una variante del diagrama estrella. En este modelo se normalizan las dimensiones creando así jerarquías y conservando la esencial tabla de hechos.
- g) **KDD:** Son las siglas en ingles del descubrimiento de conocimiento apunta a procesar automáticamente grandes cantidades de datos para encontrar conocimiento útil en ellos, es el proceso no trivial de identificar patrones validos novedosos, potencialmente útiles y comprensibles a partir de datos.

2.5 Identificación de Variables

2.5.1 Variable Independiente

X= Modelo de minería de datos

2.5.2 Variable Dependiente

Y= Servicio Académico

2.6 Definición Operativa de Variables e Indicadores

Tabla N° 01.- Operacionalización de variables

VARIABLES	DIMENCIONES	INDICADORES	
<u>Independiente</u> Modelo de minería de datos.	Escala Valorativa	Nivel de satisfacción	
<u>Dependiente</u> Y= Servicio académico	Y1= Condiciones básicas de Infraestructura	• La Comodidad del mobiliario es adecuada para prestar un correcto servicio académico • Los espacios para la enseñanza (aulas, laboratorios), son apropiados para prestar un correcto servicio académico • Los espacios para el estudio, se distribuye apropiadamente para prestar un correcto servicio académico • Las Instalaciones sanitarias son adecuadas para prestar un correcto servicio académico • Las disponibilidades de agua potable son adecuadas para prestar un correcto servicio académico. • La biblioteca es adecuadas para prestar un correcto servicio académico.	Ítem del 1 al 6
	Y2= Proceso enseñanza aprendizaje	• El proceso de matrícula, son adecuados para prestar un correcto servicio académico. • Los materiales impresos y audiovisuales disponibles son adecuados para prestar un correcto servicio académico. • El dominio de los contenidos y actualización de los profesores son adecuados para prestar un correcto servicio académico. • El sistema de evaluación es adecuado para prestar un correcto servicio académico • La metodología de enseñanza es adecuada para prestar un correcto servicio académico • La orientación y apoyo de los profesores en las horas de consulta son adecuados para prestar un correcto servicio académico • La incorporación de las nuevas tecnologías a la enseñanza aprendizaje del estudiante son adecuados para prestar un correcto servicio académico.	Ítem del 7 al 13

Fuente: Propia

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1 Ámbito de estudio

El estudio se desarrollará en la Institución Educativa Harvard.

3.2 Tipo de Investigación

Según su propósito **la investigación es aplicada**, al respecto; (Murillo, 2008), la investigación aplicada recibe el nombre de “investigación práctica o empírica”, que se caracteriza porque busca la aplicación o utilización de los conocimientos adquiridos, a la vez que se adquieren otros, después de implementar y sistematizar la práctica basada en investigación.

3.3 Nivel de Investigación

El nivel de investigación es descriptivo – explicativo, es descriptiva ya que esta investigación nos dará información de la situación de **la Institución Educativa Harvard** tal y como se muestra actualmente como por ejemplo cual es el giro del negocio, como están organizados, características internas, externas de sus propiedades, algunas dificultades que esta presentan, etc. Todo esto en un momento y tiempo histórico concreto y determinado. Por otro lado, es explicativa ya nos centraremos en descubrir que variables se relacionan y como se relacionan dando a conocer las causas o factores que han dado origen o han condicionado la existencia y naturaleza del hecho o fenómeno de estudio.

3.4 Método de Investigación

Como método general se utilizó el método científico.

Cómo método específico se utilizó el método de la observación y del análisis síntesis.

3.5 Diseño de la Investigación

Dada la naturaleza del problema y el objetivo de investigación, en el proyecto de investigación se utilizará el diseño de investigación Cuasi-experimental con Pre test y Post test con grupo experimental, cuyo esquema es el siguiente:

G. E.: O₁ X O₂

Dónde:

G.E.: Grupo experimental

O₁: Pre Test, es la prueba inicial

O₂: Post Test, es la post prueba

X: Manipulación de la Variable Independiente

3.6 Población, Muestra y Muestreo

a. **Población:** Está compuesto por los consta de 250 alumnos.

TRABAJADORES
250

b. **Muestra:** La muestra es la que puede determinar la problemática ya que le es capaz de generar los datos con los cuales se identifican las fallas dentro del proceso.

- Muestra es la siguiente:

$$n_0 = \frac{z^2 s^2}{d^2}$$

n_0 = tamaño de la muestra

z = nivel de confianza

s^2 = pxq = varianza

p = muestra aproximada

q = $1 - p$

d = precision (error)

N = poblacion

$$n_0 = \frac{1.645^2 0.5^2}{0.05^2} = 187.91$$

$$n_0 = ?$$

$$z = 1.645$$

$$s^2 = pxq$$

$$p = 0.5$$

$$q = 1 - p = 0.5$$

$$d = 0.06$$

$$N = 250$$

Hallando muestra exacta:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$
$$n = \frac{187.91}{1 + \frac{187.91}{250}} = 107.27$$

$$n = 107$$

El tamaño de muestra estará constituido por 107 estudiantes, quienes aprecian la variable servicio académico.

3.7 Técnicas e Instrumentos de Recolección de Datos

3.7.1 Técnicas

Para la obtención de los datos e información en la presente investigación se utilizarán:

- Encuestas
- La observación.
- El Análisis Bibliográfico.
- Entrevistas.

3.7.2 Instrumentos.

- Cuestionarios.

3.8 Procedimiento de Recolección de Datos

Respecto del procesamiento de datos, los datos obtenidos durante el trabajo de investigación fueron procesado utilizando técnicas y métodos de la Estadística Descriptivas como: cuadros estadísticos, estadígrafos (el promedio aritmético, la mediana, la moda), los cuales son presentados a través de cuadros y gráficos estadísticos

3.9 Técnicas de Procesamiento y Análisis de Datos

Los datos obtenidos durante el trabajo de campo serán procesados utilizando técnicas y métodos de la Estadística como: cuadros estadísticos, gráficos estadísticos, estadígrafos (el promedio aritmético, la mediana, la moda), y la prueba de diferencia de proporciones como técnica para probar la hipótesis planteada.

CAPITULO IV

RESULTADOS

4.1 Presentación de Resultados

En este capítulo se presenta los resultados obtenidos en la tesis, MODELO DE MINERÍA DE DATOS PARA LA MEJORA DEL SERVICIO ACADÉMICO. Como primer resultado se muestra el modelo de minería de datos.

En la investigación se utilizó la **prueba de diferencia de proporciones**, para aceptar o rechazar las Hipótesis planteadas.

Prueba **de diferencia de proporciones**.¹

Definición: Es una prueba estadística para analizar si dos proporciones difieren significativamente entre si.

Hipótesis: De diferencia de proporciones en dos grupos.

Variable: La comparación se realiza sobre una variable.

Nivel de medición de la variable:

Intervalos o razón, expresados en proporciones o porcentajes.

Procedimiento e interpretación:

Se obtienen las proporciones de los grupos. Se aplica la siguiente fórmula:

$$Z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Dónde: $q_1 = 1 - p_1$ y $q_2 = 1 - p_2$

La puntuación “z” resultante se compara con la puntuación “z” de la distribución de puntuaciones “z”(normal) que corresponde al nivel de confianza elegido. El valor calculado de “z”(resultante de

¹ Hernández, Fernández y Baptista (2003). Metodología de la investigación. (p 387)

aplicar la formula) debe ser igual o mayor que el valor de la tabla de áreas bajo la curva normal correspondiente.

- Si es igual o Mayor, se acepta la hipótesis de Investigación.
- Si es menor, se rechaza la hipótesis de Investigación.

4.1.1 Variable Independiente “Modelo de Minería de Datos”

Los resultados de esta variable se hacen en base a la presentación del funcionamiento del Modelo de minería de datos, en base a las actividades desarrolladas en el diseño del modelo propuesto. A continuación, se presenta un resumen del procedimiento desarrollado en la construcción del modelo.

Para la construcción del modelo se utilizó la Metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

La vida útil de nuestro proyecto está dada por el avance de la tecnología y de las herramientas de inteligencia de negocio que van mejorando al transcurrir el tiempo. Es decir, nuestro proyecto tendrá una duración hasta que esta herramienta quede desplazada por una mejor.

1. BASE DE DATOS DEL COLEGIO

Código	Apellidos y Nombres	Sexo	Profesión	Experiencia	Hrs#_Capacita...	Hrs#_Tardanza	Grado_Académico	Situación
2052970	CLARO RAMOS, ...	M	EP	5	850	2	MAGISTER	Contratado
2098787	DAMIAN ARTEA...	F	EP	4	550	3	LICENCIADO	Contratado
201345	GUERRA GUERR...	M	EP	2	500	2	MAGISTER	Contratado
456723	SOTO PALACIO...	F	EP	2	250	2	LICENCIADO	Contratado
12567	HUAMAN PIÑAS...	F	ES	2	240	2	LICENCIADO	Contratado
2456678	PINEDA PIÑAS...	F	EP	1	180	1	LICENCIADO	Contratado
678923	VILCAS ARTESA...	F	EP	1	180	1	LICENCIADO	Contratado
453267	BOZA CERRON, ...	M	ES	1	172	1	LICENCIADO	Contratado
4987	CORDOVA PERE...	F	ES	2	200	2	LICENCIADO	Contratado
12398	CORDOVA HINO...	F	ES	3	350	3	MAGISTER	Contratado
9835	BELTRAN CER...	M	ES	3	340	2	MAGISTER	Contratado
87235	HUAROC LANDE...	F	EP	3	350	2	MAGISTER	Contratado
2378653	MENDEZ LOPEZ...	M	EP	4	480	2	MAGISTER	Contratado
2349876	MENDEZ ESPINO...	M	ES	4	480	2	LICENCIADO	Contratado
12987	PARJONA JULCA...	M	ES	4	490	3	LICENCIADO	Contratado
34567	PERALTA VENTU...	M	ES	5	600	3	MAGISTER	Contratado
234890	SOTELO VILLAR...	M	ES	5	590	3	MAGISTER	Contratado
235678	ALANZA TORRE...	M	EP	6	620	4	LICENCIADO	Contratado
13498	ROBLES ROBLES...	F	EP	7	680	4	MAGISTER	Contratado
234	VILLAR MONTES...	M	EP	8	820	2	MAGISTER	Contratado
236789	CHOCCA MONT...	M	EP	7	740	1	MAGISTER	Contratado
345671	TORRES TORRE...	F	ES	6	680	3	MAGISTER	Contratado
334578	POMA POMA, AL...	F	ES	5	600	4	MAGISTER	Contratado
2345	HUAMAN PEÑA, ...	M	EP	4	720	2	MAGISTER	Contratado
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Microsoft SQL Server Management Studio

Archivo Editar Ver Proyecto Depurar Diseñador de consultas Herramientas Ventana Comunidad Ayuda

Explorador de objetos

Conectar (local) (SQL Server 10.50.1600 - Bases de datos)

Instantáneas de bases de datos

Academico2

Diagramas de base de datos

Tablas

dbo.Estudiantes

dbo.Docente

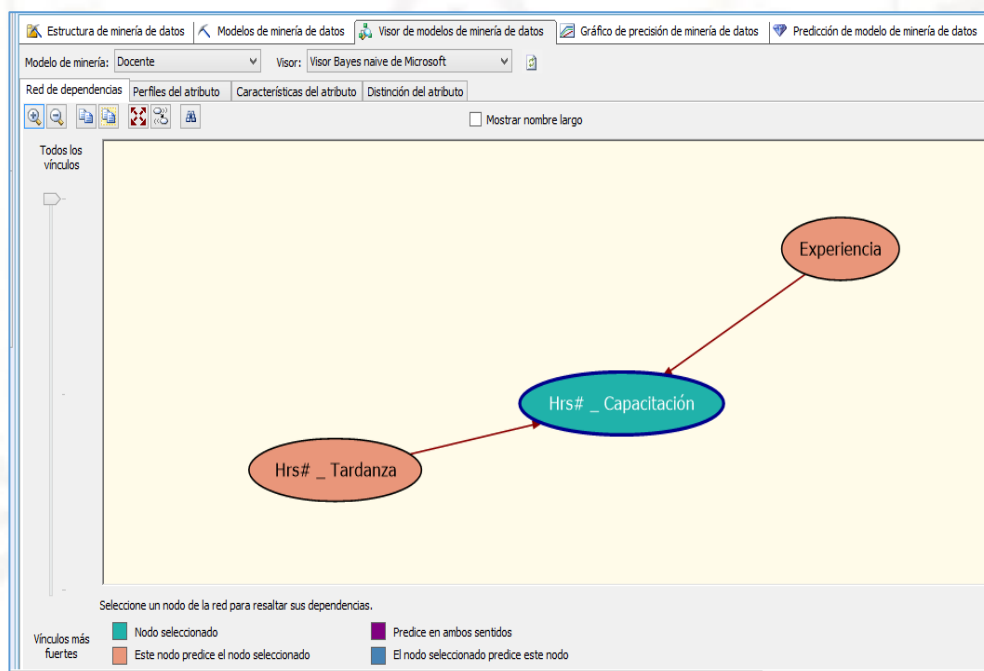
Detalles del Explorador de objetos

Código	Apellidos y Nom...	Sexo	Promedio Final	Rendimiento	Comportamiento	Procedencia	Nivel_Satisfacción	Responsable	Sexo_Respons...	Nivel de Instruc...
19876543210	ROJAS SOTO, M...	M	13	R	A	El Tambo	2	ROJAS PEÑALO...	M	SECUNDARIA
19876543210	AGUILAR LA RO...	F	15	R	B	El Tambo	2	LA ROSA SOTO...	F	SECUNDARIA
5143056789320	ARATOMA ROD...	M	16	A	AD	El Tambo	3	ARATOMA CRIS...	M	SECUNDARIA
3245672258910	MANCO SALAZA...	F	12	B	A	El Tambo	1	SALAZAR QUES...	F	SECUNDARIA
1987234561230	SOLIS LEZAMA...	F	15	R	B	Huancayo	3	SOLIS PEREZ, P...	M	PRIMARIA
41234567890	RODRIGUEZ ME...	F	16	A	A	Huancayo	2	RODRIGUEZ SO...	M	UNIVERSITARIA
2156834389090	HUAROTO TUNIC...	F	14	R	B	Huancayo	2	TUNICAR DIEGO...	F	UNIVERSITARIA
2315689232340	LEDESMA CURO...	M	14	R	B	Huancayo	1	CURO RAMOS, ...	F	SECUNDARIA
2633485739080	LEDESMA SALAZ...	F	16	A	A	Huancayo	1	SALAZAR PEÑA...	F	UNIVERSITARIA
4534456788230	RODRIGUEZ SO...	F	11	B	C	Huancayo	1	SOTO RIMAC, M...	F	PRIMARIA
4356789531210	SOTO MACUJIC...	M	17	A	AD	Huancayo	2	SOTO PALOMAR...	M	SECUNDARIA
2334133435450	DAMIAN SOLIS...	F	14	R	B	Huancayo	1	DAMIAN ARTEA...	M	UNIVERSITARIA
2334345556560	GUERRA SOLIS...	M	12	B	C	Huancayo	3	GUERRA GUERR...	M	UNIVERSITARIA
5667878234380	DANTE CORNEJ...	M	13	R	C	Huancayo	2	DANTE GUZMAN...	M	SECUNDARIA
456788734350	CORNEJO SALA...	M	12	B	C	Huancayo	1	SALAZAR POMA...	F	UNIVERSITARIA
1223233454570	CORNEJO ROJA...	M	12	B	C	Chupaca	3	CORNEJO AVELL...	M	UNIVERSITARIA
3445790213430	CURO SOTO, LU...	M	14	R	B	Chupaca	2	SOTO PALACIO...	F	SECUNDARIA
4556734567800	HUAMANI PALA...	M	13	R	A	Chupaca	2	HUAMANI PIÑAS...	M	UNIVERSITARIA
2233546567890	ROJAS SOVERO...	F	15	R	A	Chupaca	3	SOVERO PILAR...	F	SECUNDARIA
3445760178990	SALAZAR OBIED...	M	16	A	A	Chupaca	1	SALAZAR MITM...	M	UNIVERSITARIA
4567832341231	ROJAS SARMIE...	M	14	R	A	Chupaca	1	ROJAS HUAMAN...	M	UNIVERSITARIA
97011856479930	ARROYO RAFAE...	F	13	R	A	Cajas	2	ARROYO PEÑA...	M	UNIVERSITARIA
NULL	ARTEZANO PIN...	F	15	R	A	Cajas	1	PINEDA PIÑAS...	F	UNIVERSITARIA
NULL	BAUTISTA VIL...	M	14	R	A	Cajas	2	VILCAS ARTEA...	F	UNIVERSITARIA
NULL	CAPCHA AGUIR...	M	13	R	A	Cajas	2	CAPCHA PIÑAS...	F	UNIVERSITARIA

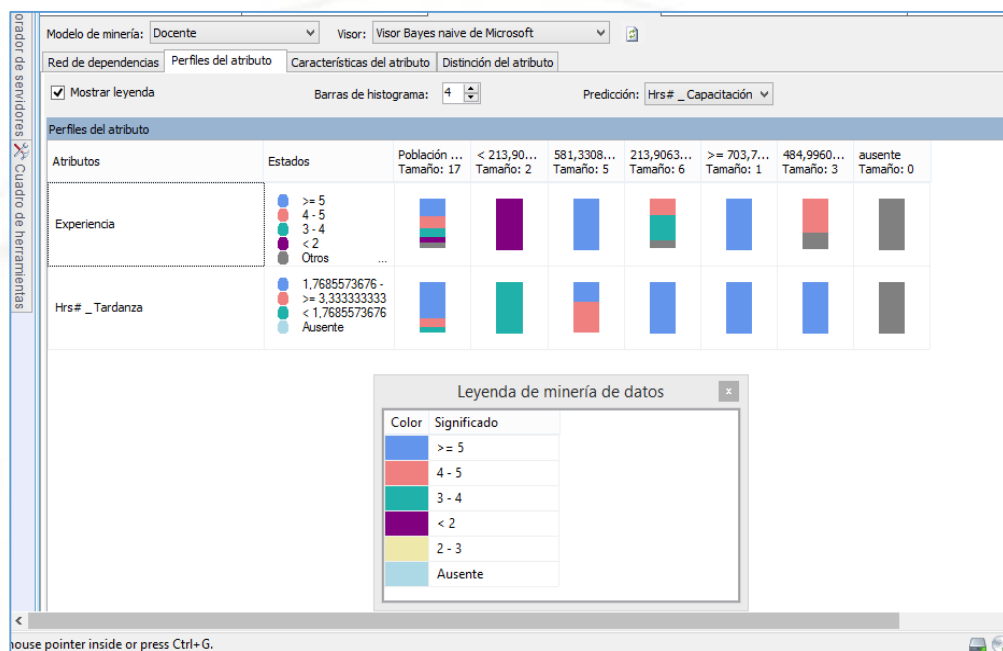
21:07 24/08/2017

2. MINERIA DE DATOS DOCENTE: VISOR BAYES NAIVE

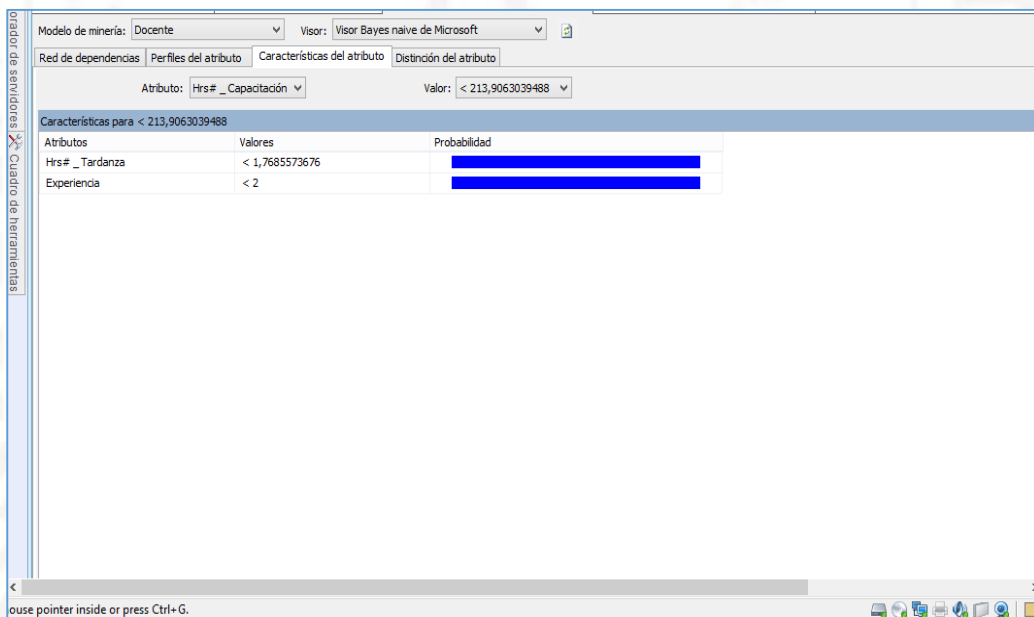
RED DE DEPENDENCIAS



PERFILES DE ATRIBUTO



CARACTERÍSTICAS DEL ATRIBUTO



DISTINCIÓN DEL ATRIBUTO

Modelo de minería: Docente Visor: Visor Bayes naïve de Microsoft

Red de dependencias Perfiles del atributo Características del atributo Distinción del atributo

Atributo: Hrs#_Capacitación Valor 1: < 213,9063039488 Valor 2: Todos los otros esta

Puntuaciones de distinción para < 213,9063039488 y Todos los otros estados

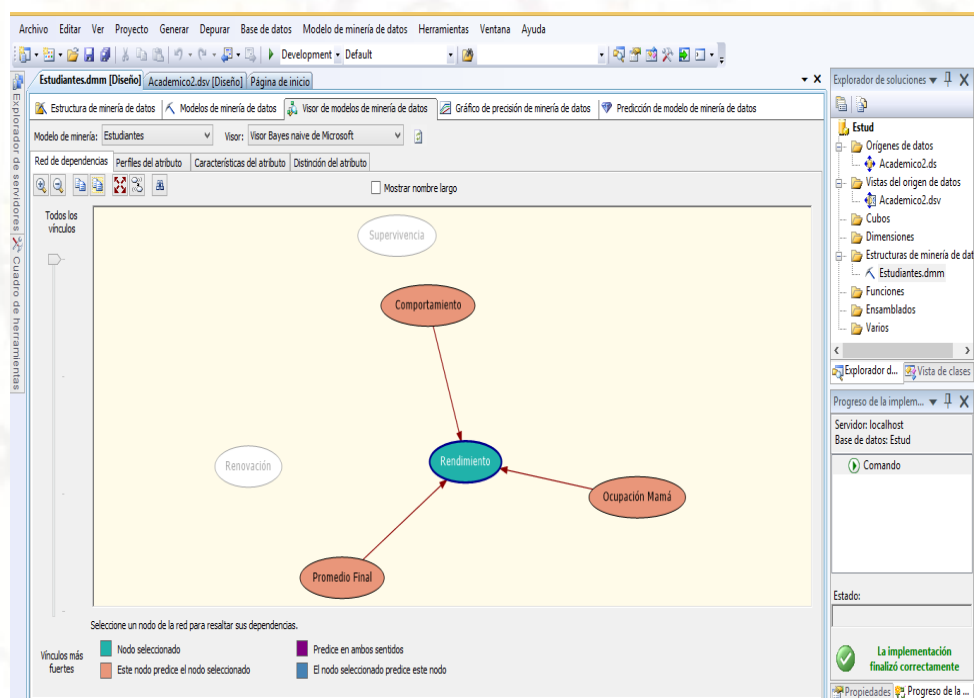
Atributos	Valores	Favorece < 213,9063039...	Favorece Todos los otros ...
Hrs#_Tardanza	< 1,7685573676		
Experiencia	< 2		
Hrs#_Tardanza	1,7685573676 - 3,333333332		

Leyenda de minería de datos

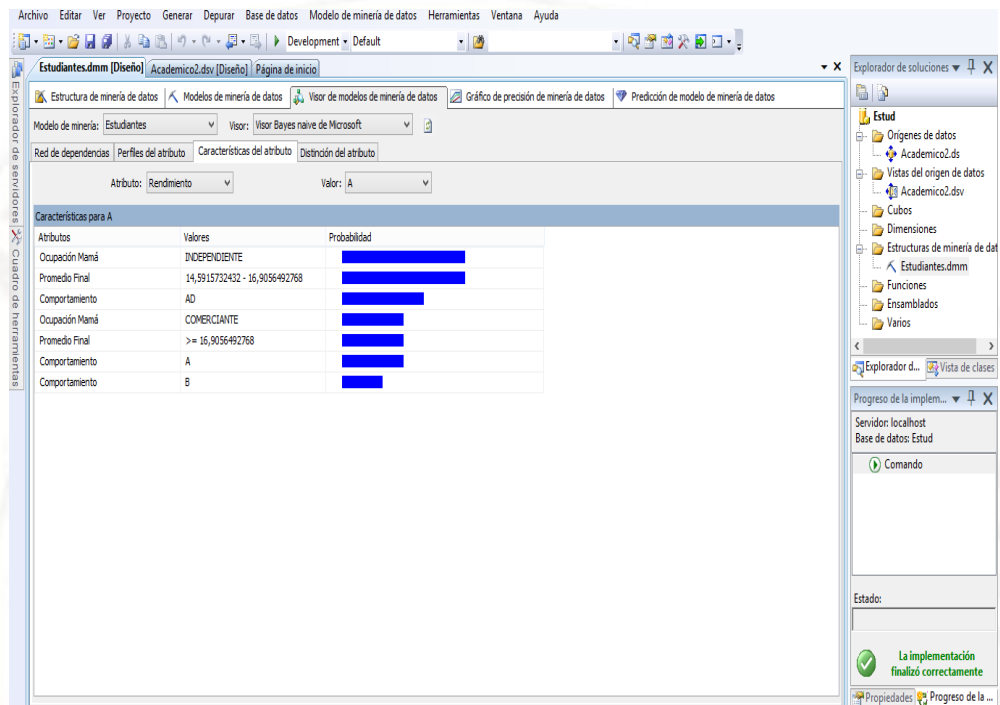
	< 213,9063039488	Todos los otros
Hrs#_Tardanza =...	2	0
Hrs#_Tardanza !...	0	15

3. MODELO DE MINERÍA DE DATOS DE ESTUDIANTES: VISOR BAYES NAIVES

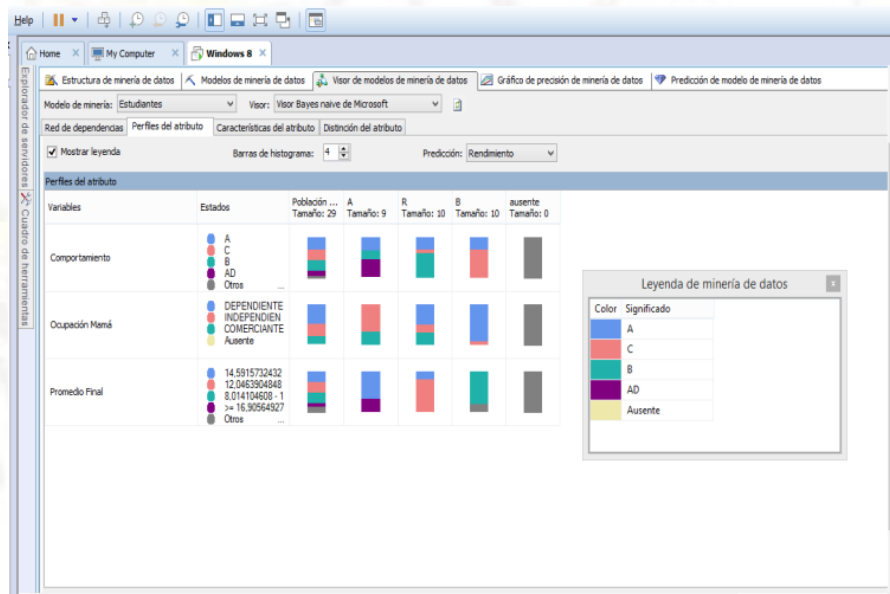
RED DE DEPENDENCIAS



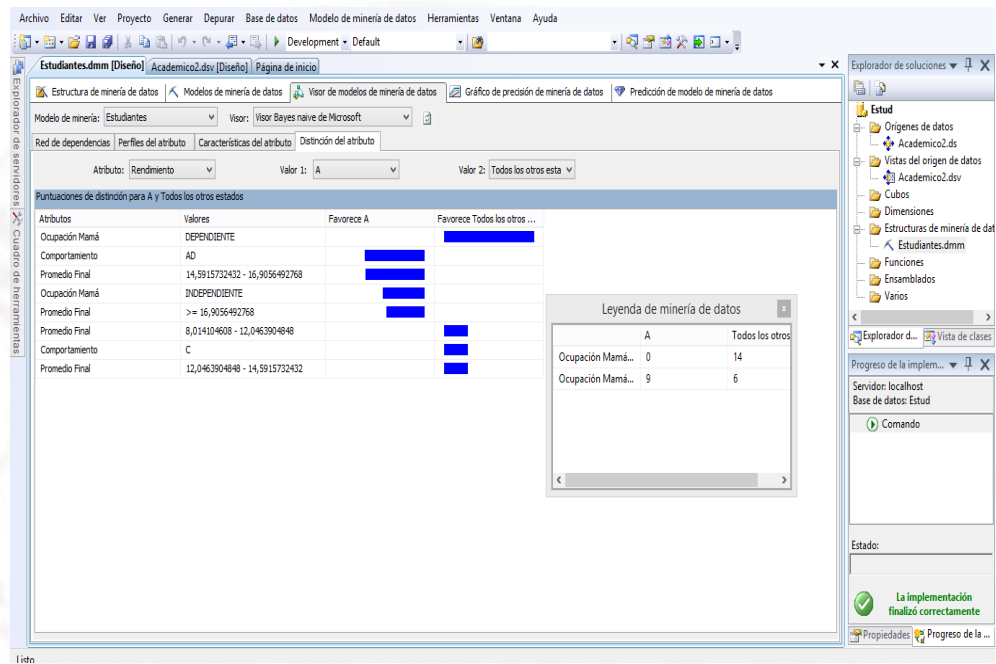
PERFILES DE ATRIBUTO



CARACTERISTICAS DEL ATRIBUTO



DISTINCIÓN DEL ATRIBUTO



4.1.2 Variable Dependiente “Servicio Académico”

Para mostrar los resultados de la prueba de hipótesis, se utilizó la **prueba de diferencia de proporciones**, para aceptar o rechazar las Hipótesis planteadas.

A **continuación**, presentamos los resultados de la variable dependiente.

RESULTADOS OBTENIDOS: MODELO DE MINERIA DE DATOS-CONDICIONES BASICAS DE INFRAESTRUCTURA, EN LA PRE PRUEBA Y POST PRUEBA

Tabla N^a 02. Resultados obtenidos: Modelo de Minería de datos-Condiciones básicas de infraestructura

PREGUNTAS	C CATEG.	GRUPO EXPERIMENTAL			
		PRE PRUEBA		POST PRUEBA	
		fi	%	fi	%
La Comodidad del mobiliario es adecuada para prestar un correcto servicio académico	SI	26	26.92%	101	94.23%
	NO	79	73.08%	6	5.77%
	Σ	107	100.00%	107	100.00%
Los espacios para la enseñanza (aulas, laboratorios), son apropiados para prestar un correcto servicio académico	SI	22	19.23%	104	97.12%
	NO	85	80.77%	3	2.88%
	Σ	107	100.00%	107	100.00%
Los espacios para el estudio, se distribuye apropiadamente para prestar un correcto servicio académico	SI	12	11.54%	101	94.23%
	NO	95	88.46%	6	5.77%
	Σ	107	100.00%	107	100.00%
Las Instalaciones sanitarias son adecuadas para prestar un correcto servicio académico	SI	12	11.54%	104	97.12%
	NO	95	88.46%	3	2.88%
	Σ	107	100.00%	107	100.00%
Las disponibilidades de agua potable son adecuadas para prestar un correcto servicio académico.	SI	11	10.58%	98	91.35%
	NO	96	89.42%	9	8.65%
	Σ	107	100.00%	107	100.00%
La biblioteca es adecuadas para prestar un correcto servicio académico.	SI	14	13.46%	95	86.54%
	NO	93	86.54%	12	13.46%
	Σ	107	100.00%	107	100.00%

FUENTE: Elaboración propia

MODELO DE MINERÍA DE DATOS-CONDICIONES BÁSICAS DE INFRAESTRUCTURA

Tabla N° 03. Modelo de Minería de datos-Condiciones básicas de infraestructura

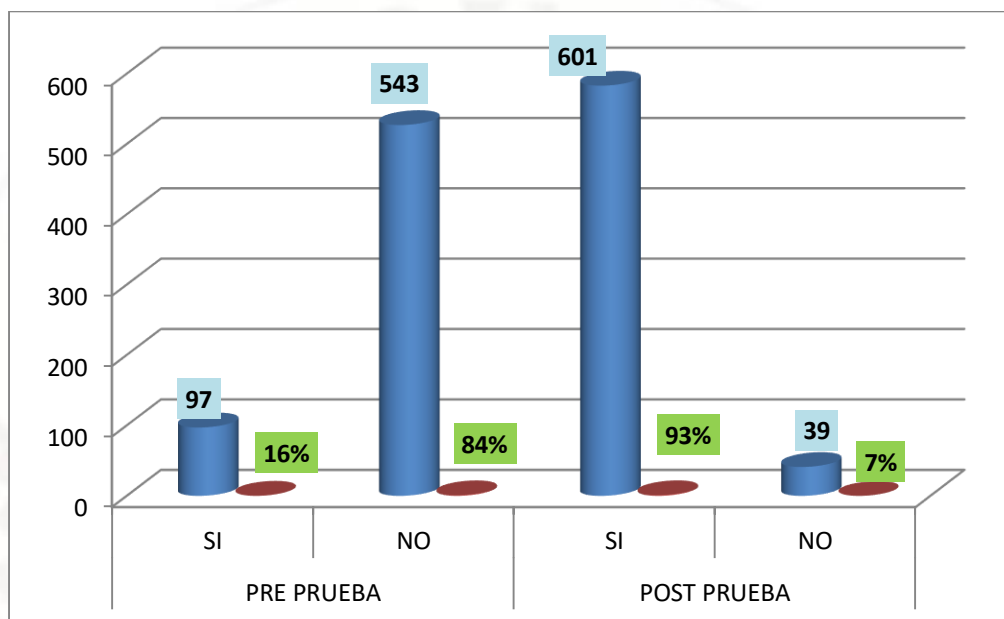
PREG.	GRUPO EXPERIMENTAL							
	PRE PRUEBA				POST PRUEBA			
	SI		NO		SI		NO	
	fi	%	fi	%	fi	%	fi	%
01	26	26.92%	79	73.08%	101	94.23%	6	5.77%
02	22	19.23%	85	80.77%	104	97.12%	3	2.88%
03	12	11.54%	95	88.46%	101	94.23%	6	5.77%
04	12	11.54%	95	88.46%	104	97.12%	3	2.88%
05	11	10.58%	96	89.42%	98	91.35%	9	8.65%
06	14	13.46%	93	86.54%	93	86.54%	12	13.46%
Σ	97	93.27%	543	507%	601	561%	39	39%
P	0.16		0.84		0.93		0.07	

Fuente: Propia

Análisis e Interpretación:

97 respuestas que representa el 16% correspondiente a la dimensión de SI del pre prueba y 561 respuestas que representan el 93% correspondiente a la dimensión de SI de la post prueba; demuestran claramente que la aplicación del Modelo de minería de datos SI influye en la mejora de las condiciones básicas de infraestructura en la I.E.P. HARVARD – HUANCAYO

Gráfico N° 21. Modelo de Minería de datos-Condiciones básicas de infraestructura



FUENTE: Propia

Prueba de la Hipótesis Especifica n° 01.

n modelo de minería de datos influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo.

1. Planteamiento de las Hipótesis:

Ho: P post prueba = P pre prueba.

Ha: P post prueba > P pre prueba.

2. Determinación del Nivel de Significancia:

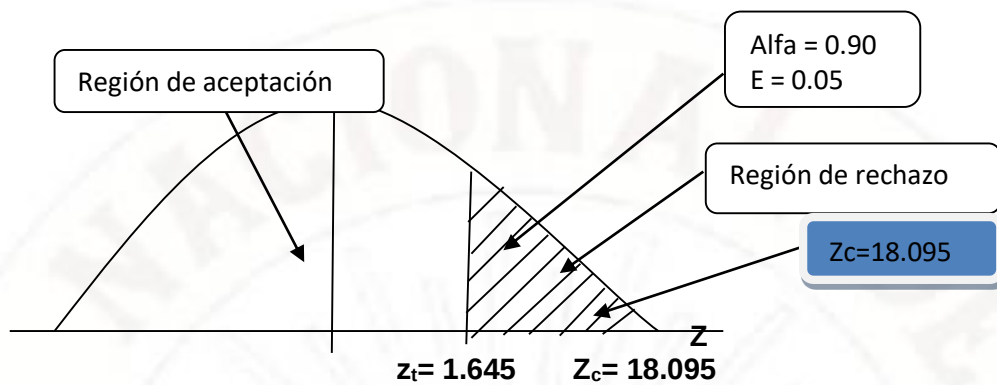
Alfa = 90%

Error= 0.05

3. Estadístico de Prueba:

$$Z = \frac{(P_2 - P_1)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

4. Determinación de las Regiones Críticas:



5. Cálculo del Valor Tabular:

$$T_t = 1.645$$

6. Cálculo del Valor Experimental:

$$Z = \frac{(P_2 - P_1)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

DATOS:

$$P_1 = 97/640 = 0,16$$

$$P_2 = 601/640 = 0,93$$

$$q_1 = 0,84$$

$$q_2 = 0,07$$

$$n_1 = 107$$

$$n_2 = 107$$

Reemplazando valores en la fórmula, tenemos:

$$Z = \frac{(0.93 - 0.16)}{\sqrt{\frac{(0.16)(0.84)}{107} + \frac{(0.93)(0.07)}{107}}}$$
$$Z = 18.095$$

7. Decisión Estadística:

Como $Z_c = 18.095 > Z_t = 1.645$ y $H_0 \in RR$, entonces se rechaza la hipótesis H_0 y se acepta la hipótesis H_a , en este sentido se afirma tajantemente que la P de la post prueba es mayor que la P de pre prueba. Con ello afirmamos que Un modelo de minería de datos influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo.

RESULTADOS OBTENIDOS: MODELO DE MINERIA DE DATOS- PROCESO ENSEÑANZA APRENDIZAJE, EN LA PRE PRUEBA Y POST PRUEBA

Tabla N° 04. Resultados obtenidos Modelo de Minería de datos - Proceso enseñanza aprendizaje.

PREGUNTAS	C	GRUPO EXPERIMENTAL			
	CATEG	PRE PRUEBA		POST PRUEBA	
		fi	%	fi	%
El proceso de matrícula, son adecuados para prestar un correcto servicio académico.	SI	3	2.88%	102	95.19%
	NO	104	97.12%	5	4.81%
	Σ	107	100.00%	107	100.00%
Los materiales impresos y audiovisuales disponibles son adecuados para prestar un correcto servicio académico.	SI	6	5.77%	103	96.15%
	NO	101	94.23%	4	3.85%
	Σ	107	100.00%	107	100.00%
El dominio de los contenidos y actualización de los profesores son adecuados para prestar un correcto servicio académico.	SI	10	9.62%	104	97.12%
	NO	97	90.38%	3	2.88%
	Σ	107	100.00%	107	100.00%
El sistema de evaluación es adecuado para prestar un correcto servicio académico	SI	8	7.69%	101	94.23%
	NO	99	92.31%	6	5.77%
	Σ	107	100.00%	107	100.00%
La metodología de enseñanza es adecuado para prestar un correcto servicio académico	SI	11	10.58%	103	96.15%
	NO	96	89.42%	4	3.85%
	Σ	107	100.00%	107	100.00%
La orientación y apoyo de los profesores en las horas de consulta son adecuados para prestar un correcto servicio académico	SI	9	8.65%	105	98.08%
	NO	98	91.35%	2	1.92%
	Σ	107	100.00%	107	100.00%

7a incorporación de las nuevas tecnologías a la enseñanza aprendizaje del estudiante son adecuados para prestar un correcto servicio académico	SI	8	7.69%	101	94.23%
	NO	99	92.31%	6	5.77%
	Σ	107	100.00%	107	100.00%

Fuente: Propia

MODELO DE MINERIA DE DATOS- PROCESO ENSEÑANZA APRENDIZAJE

Tabla N° 05. Resumen Modelo de Minería de datos - Proceso enseñanza aprendizaje.

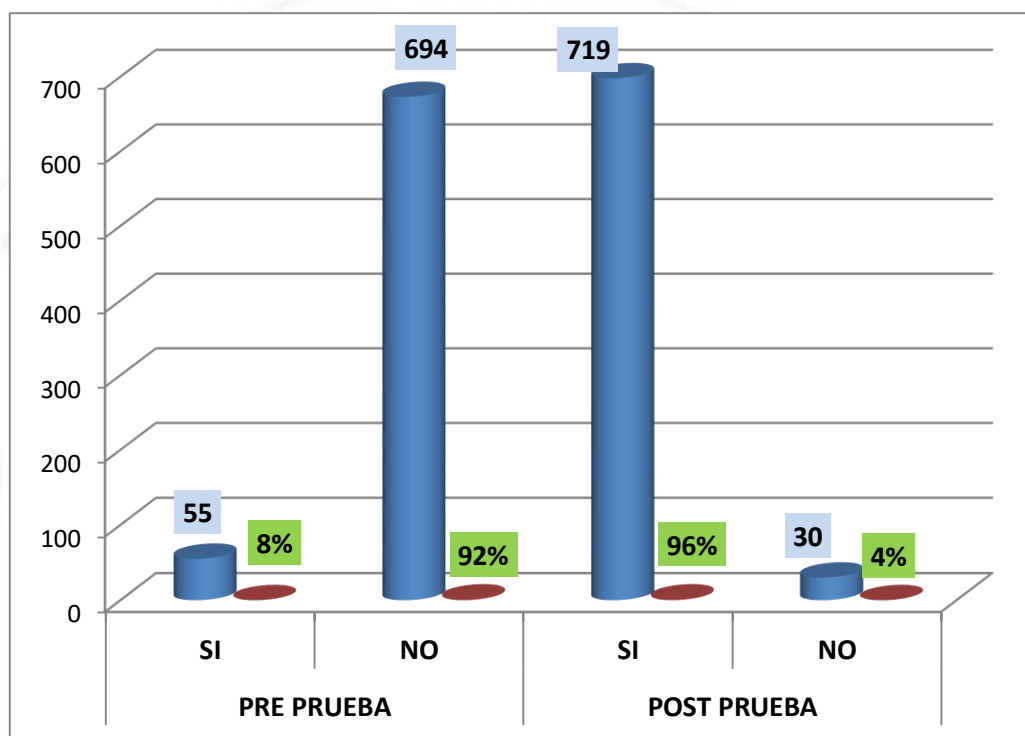
PREG.	GRUPO EXPERIMENTAL							
	PRE PRUEBA				POST PRUEBA			
	SI		NO		SI		NO	
	fi	%	fi	%	fi	%	fi	%
01	3	2.88%	104	97.12%	102	95.19%	5	4.81%
02	6	5.77%	101	94.23%	103	96.15%	4	3.85%
03	10	9.62%	97	90.38%	104	97.12%	3	2.88%
04	8	7.69%	99	92.31%	101	94.23%	6	5.77%
05	11	10.58%	96	89.42%	103	96.15%	4	3.85%
06	9	8.65%	98	91.35%	105	98.08%	2	1.92%
07	8	7.69%	99	92.31%	101	94.23%	6	5.77%
Σ	55	52.88%	694	647.12%	719	671.15%	30	28.85%
P	0.08		0.92		0.96		0.04	

Fuente: Propia

Análisis e Interpretación:

55 respuestas que representa el 8% correspondiente a la dimensión de SI de la pre prueba y 719 respuestas que representan el 96% correspondiente a la dimensión de SI de la post prueba; demuestran claramente que un modelo de minería de datos si influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo.

Gráfico N° 22. Resumen Modelo de Minería de datos - Proceso enseñanza aprendizaje.



Fuente: Propia

a) Prueba de la Hipótesis Especifica N° 02.

Un modelo de minería de datos SI influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo

1. Planteamiento de las Hipótesis:

$H_0: P \text{ post prueba} = P \text{ pre prueba}.$

$H_a: P \text{ post prueba} > P \text{ pre prueba}.$

2. Determinación del Nivel de Significancia:

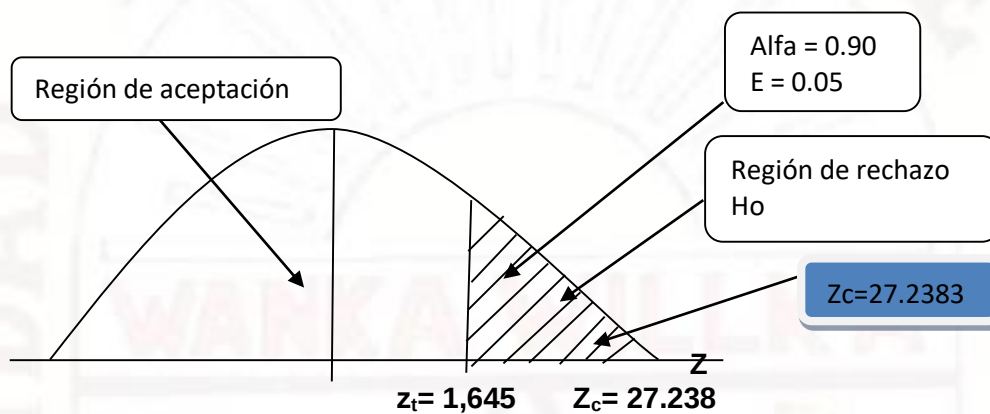
Alfa = 90%

Error = 0,05

3. Estadístico de Prueba:

$$Z = \frac{(P_2 - P_1)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Determinación de las Regiones Críticas:



4. Cálculo del Valor Tabular:

$$T_t = 1,645$$

5. Cálculo del Valor Experimental:

$$Z = \frac{(P_2 - P_1)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

DATOS:

$$P_1 = 55/749 = 0,08$$

$$P_2 = 719/749 = 0,96$$

$$q_1 = 0,92$$

$$q_2 = 0,04$$

$$n_1 = 104$$

$$n_2 = 104$$

Reemplazando valores en la fórmula, tenemos:

$$Z = \frac{(0,96 - 0,08)}{\sqrt{\frac{(0,08)(0,92)}{107} + \frac{(0,96)(0,04)}{107}}}$$

$$Z = 27.238$$

6. Decisión Estadística:

Como $Z_c = 27.238 > Z_t = 1,645$ y $H_0 \in RR$, entonces se rechaza la Hipótesis H_0 y se acepta la hipótesis H_a , en este sentido se afirma tajantemente que la P de la post prueba es mayor que la P de la pre prueba. Con ello afirmamos que el modelo de minería de datos SI influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo.

4.2 Discusión

Del resultado de análisis de la variable de estudio se obtiene el resultado de contrastación empírica de la hipótesis planteada.

De los entrevistados de la muestra consideran que es necesario en el proceso de mejorar el servicio académico implementar un modelo de minería de datos que ayude en este proceso en la I.E.P. Harvard., entendiéndose que la variable servicio académico se midió en sus dos dimensiones¹ condiciones básicas de infraestructura, 2 Proceso enseñanza aprendizaje.

Con los resultados obtenidos podemos corroborarlo al encontrar en la prueba de hipótesis específica 1. 97 respuestas que representa el 16% a la dimensión de SI de la pre prueba y 561 respuestas que representan el 93% a la dimensión de SI de la post prueba; De otro lado al analizar los resultados obtenidos, en la prueba de la Z, podemos observar que $Z_C = 18.095 > Z_t = 1.645$ y $H_0 \in RR$,. Con ello afirmamos que Un modelo de minería de datos SI influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo.

Respecto a la prueba de hipótesis específica 2. Se puede observar que 55 respuestas que representa el 8% a la dimensión de SI de la pre prueba y 719 respuestas que representan el 96% a la dimensión de SI de la post prueba; demuestran claramente que un modelo de minería de datos si influye en la mejora del Proceso enseñanza aprendizaje en la Institución

Educativa Particular Harvard-Huancayo, esto se corrobora que al realizar la prueba de la “Z” se obtuvo que $ZC = 27.238 > Zt = 1,645$ y $H_0 \in RR$, Con ello se reafirma que el modelo de minería de datos SI influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo.

Comprobado las dos hipótesis específicas podemos concluir que el modelo de minería de datos SI influye en la mejora del Servicio Académico en la Institución Educativa Particular Harvard-Huancayo.

CONCLUSIONES

1. El desarrollo de la implementación de un modelo de minería de datos permitió determinar la herramienta adecuada para mejorar el servicio académico de la Institución Educativa Harvard.
2. De la investigación se concluye que un modelo de minería de datos SI influye en la mejora del Servicio Académico en la Institución Educativa Particular Harvard-Huancayo, esto se confirma al haber comprobado que el modelo de minería de datos influye en las dimensiones de “condiciones básicas de Infraestructura” y en el “proceso de enseñanza aprendizaje”.
3. Un modelo de minería de datos SI influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo. Se evidencia al corroborarlo que 561 respuestas que representan el 93% a la dimensión de SI de la post prueba respecto de 97 respuestas que representa el 16% a la dimensión de SI del pre prueba y **al observar el valor de la prueba de la Z**, en donde $Z_C = 18.095 > Z_t = 1.645$ y $H_0 \in RR$, confirmamos la influencia del Modelo de Minería de datos.
4. Se afirma que un modelo de minería de datos SI influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo. Se evidencia al corroborar que 719 respuestas que representan el 96% a la dimensión de SI de la post prueba respecto de 55 respuestas que representa el 8% de SI del pre prueba y **al observar el valor de la prueba de la Z**, en donde $Z_C = 18.095 > Z_t = 1.645$ y $H_0 \in RR$, confirmamos la influencia del Modelo de Minería de datos.

RECOMENDACIONES

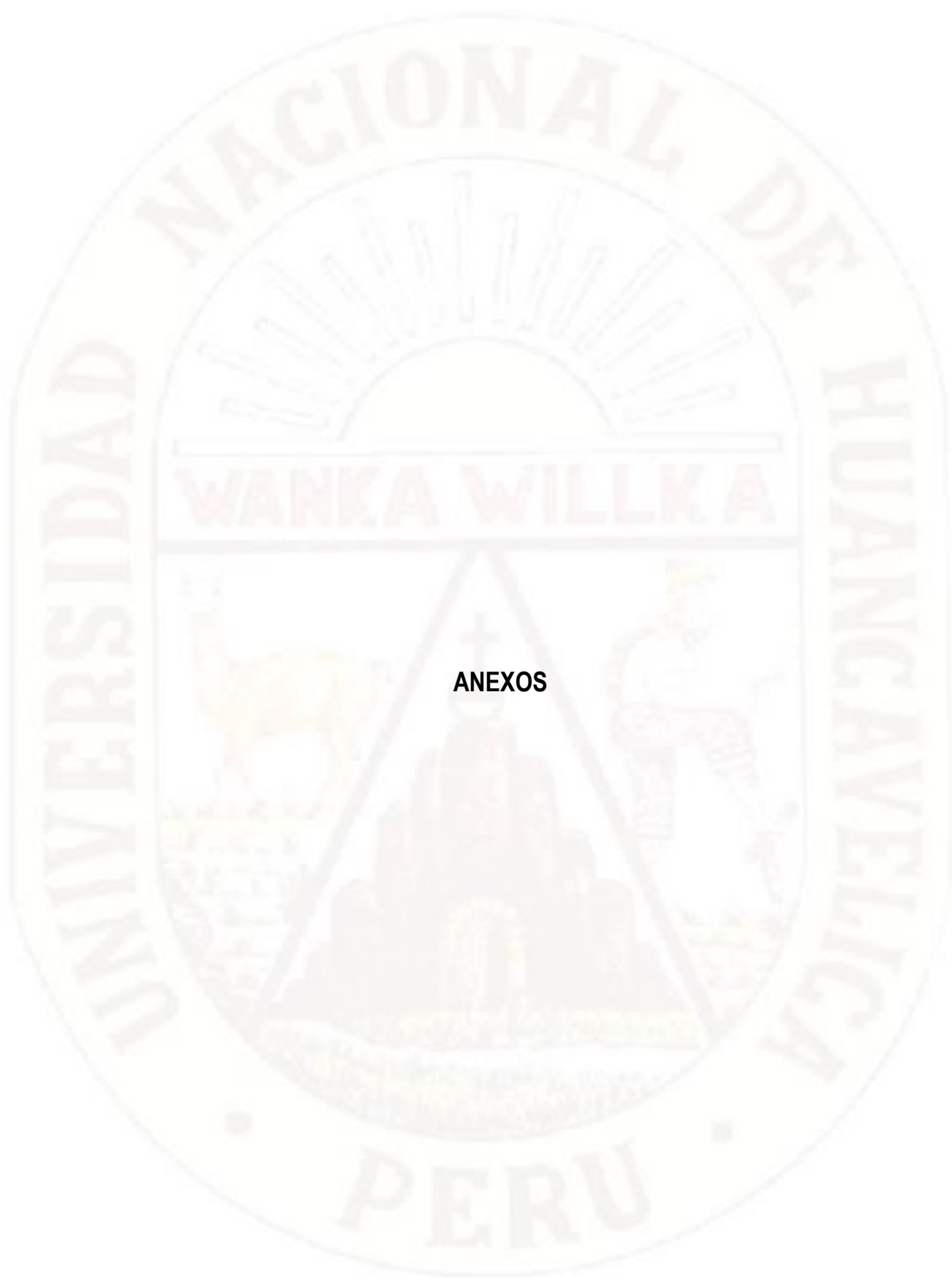
1. Se sugiere capacitar a los usuarios sobre las posibilidades del uso de Excel como herramienta específica para la explotación de datos de minería de datos.
2. Se recomienda identificar otras áreas críticas de la institución de manera de integrar toda la información estratégica de la Institución Educativa Harvard en una solución de inteligencia de negocios.
3. Se sugiere definir apropiadamente los datos en el proceso de análisis con el modelo de minería de datos.
4. Se recomienda que se sensibilice a docentes, personal administrativo y estudiantes del colegio, la importancia de la minería de datos.

REFERENCIA BIBLIOGRÁFICA

1. McLeod, R. (2001). Sistemas de información Gerencial - 658.011. Pearson Education.
2. Kenneth, L. C. (2004). Sistemas de Información Gerencial – Administración de la empresa digital, 658.04/L29. Pearson Education.
3. J.S.Hammond, P. . (2003). La Toma de Decisiones. Ediciones Deusto S.A.
Harold Koontz Y Heinz Weihrich “Administración una Perspectiva Global” mcgraw- hill interamericana de editores, s.a., 11va edición, 1999.
4. Lardent, Alberto, R. “Sistemas de información para la gestión empresarial: planeamiento, tecnología y calidad”. 1era. Edición. Editorial. Pearson. Buenos Aires. 2001.
5. Ralph Kimball.1996“. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses”. 3era edición. JohnWiley&Sons. ISBN 0-471-15337-0.
6. C.j. Date: “Introducción a los sistemas de bases de datos”. 1er. Edición Pearson Educación, 2001. ISBN 968-444-419.
7. Espinoza, Ciro.”Metodología de investigación tecnológica”. Pensando en sistemas. 1ra edición. Huancayo, Perú: Imagen Gráfica, marzo de 2010.

PÁGINAS WEB

8. “BI Solutions”. Disponible en: <http://www.oracle.com> Consulta: 20 marzo de 2011.
9. Bitam. “Business Intelligence”. [en línea]. Disponible en: <http://www.bitam.com/spanish/AcercaDeBI.htm> Consulta: 23 marzo del2011.
10. Microsoft Corporation. “Datawarehousing Training”. [en línea]. Disponible en: <http://www.microsoftsqlserver.com> . Consulta 22 de marzo del 2011.
11. Claudio Cesares. Data Warehousing. Disponible en <http://personal.lobocom.es/claudio/gen006.htm>. consulta: 22 marzo del 2011.Maribel Sabana Mendoza, Analysis Services 2008,2008 – Publicado por: Editorial Megabyte, Perú.
12. Xavier Mendoza,Business Intelligence competir con Información ,Primera Edición, 2010 – publicado por : Prentice Hall, España



ANEXOS

Anexo N° 01. Matriz de Consistencia

MODELO DE MINERÍA DE DATOS PARA LA MEJORA DEL SERVICIO ACADÉMICO EN LA INSTITUCION EDUCATIVA PARTICULAR HARVARD – HUANCAYO

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPOTESIS GENERAL	VARIABLES	DIMENCIONES	INDICADORES	
¿Podría un modelo de minería de datos influir en la mejora del servicio académico en la Institución Educativa Particular Harvard-Huancayo?	Determinar si un modelo de minería de datos podría influir en la mejora del servicio académico en la Institución Educativa Particular Harvard-Huancayo	Un modelo de minería de datos si influye en la mejora del servicio académico en la Institución Educativa Particular Harvard-Huancayo.	<u>Independiente.</u> X = Modelo de minería de datos	Escala Valorativa	Nivel de satisfacción	
¿Podría un modelo de minería de datos influir en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo?	Determinar si un modelo de minería de datos podría influir en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular	Un modelo de minería de datos si influye en la mejora de las condiciones básicas de Infraestructura en la Institución Educativa Particular Harvard-Huancayo.	<u>Dependiente</u> Y= Servicio académico	Y1= Condiciones básicas de Infraestructura	• La Comodidad del mobiliario es adecuada para prestar un correcto servicio académico • Los espacios para la enseñanza (aulas, laboratorios), son apropiados para prestar un correcto servicio académico • Los espacios para el estudio, se distribuye apropiadamente para prestar un correcto servicio académico • Las Instalaciones sanitarias son adecuadas para prestar un correcto servicio académico • Las disponibilidades de agua potable son adecuadas para prestar un correcto servicio académico.	Ítem del 1 al 6

	Harvard-Huancayo				• La biblioteca es adecuada para prestar un correcto servicio académico.	
¿Podría un modelo de minería de datos influir en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo?	Determinar si un modelo de minería de datos podría influir en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo	Un modelo de minería de datos si influye en la mejora del Proceso enseñanza aprendizaje en la Institución Educativa Particular Harvard-Huancayo.		Y2= Proceso enseñanza aprendizaje	• El proceso de matrícula, son adecuados para prestar un correcto servicio académico. • Los materiales impresos y audiovisuales disponibles son adecuados para prestar un correcto servicio académico. • El dominio de los contenidos y actualización de los profesores son adecuados para prestar un correcto servicio académico. • El sistema de evaluación es adecuado para prestar un correcto servicio académico • La metodología de enseñanza es adecuada para prestar un correcto servicio académico • La orientación y apoyo de los profesores en las horas de consulta son adecuados para prestar un correcto servicio académico • La incorporación de las nuevas tecnologías a la enseñanza aprendizaje del estudiante son adecuados para prestar un correcto servicio académico.	Ítem del 7 al 13

Anexo N° 02 Cuestionario Percepción del Servicio Académico

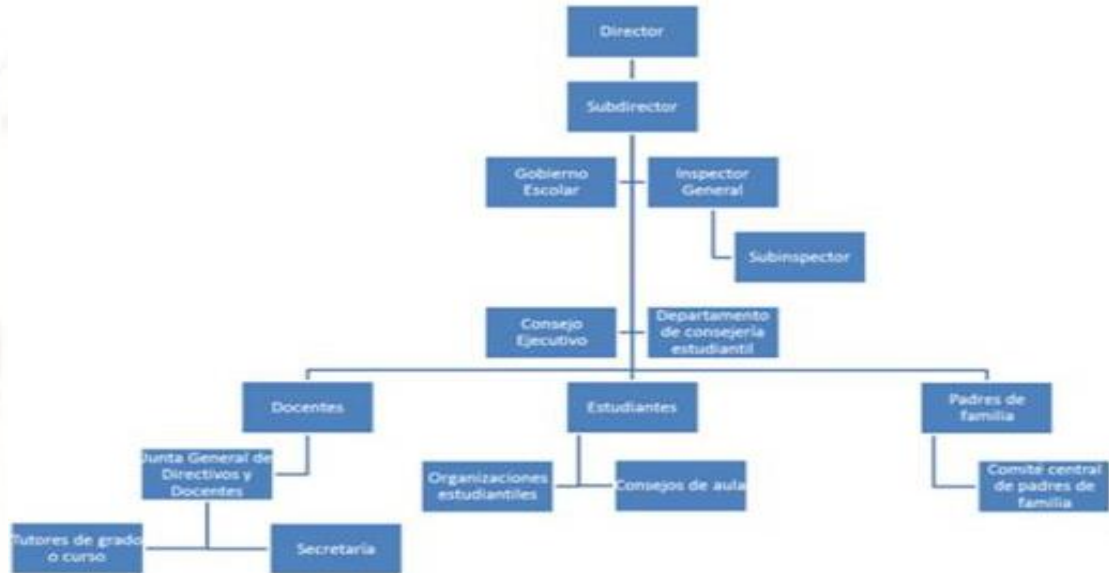
Instrucciones: El cuestionario que se le presenta a continuación es personal, se le pide por favor responder en forma anónima.

Su colaboración y sinceridad al responder las preguntas es de suma importancia. Marque con un aspa (X) su respuesta.

Ítem	PREGUNTAS	Escala	
		SI=1	NO=2
Condiciones básicas de Infraestructura			
1	La Comodidad del mobiliario es adecuada para prestar un correcto servicio académico		
2	Los eespacios para la enseñanza (aulas, laboratorios), son apropiados para prestar un correcto servicio académico		
3	Los espacios para el estudio, se distribuye apropiadamente para prestar un correcto servicio académico		
4	Las Instalaciones sanitarias son adecuadas para prestar un correcto servicio académico		
5	La disponibilidad de agua potable son adecuadas para prestar un correcto servicio académico.		
6	La biblioteca es adecuadas para prestar un correcto servicio académico.		
Proceso enseñanza aprendizaje			
7	El proceso de matrícula, son adecuados para prestar un correcto servicio académico.		
8	Los materiales impresos y audiovisuales disponibles son adecuados para prestar un correcto servicio académico.		
9	El dominio de los contenidos y actualización de los profesores son adecuados para prestar un correcto servicio académico.		
10	El sistema de evaluación es adecuado para prestar un correcto servicio académico		
11	La metodología de enseñanza es adecuado para prestar un correcto servicio académico		
12	La orientación y apoyo de los profesores en las horas de consulta son adecuados para prestar un correcto servicio académico		
13	La iincorporación de las nuevas tecnologías a la enseñanza aprendizaje del estudiante son adecuados para prestar un correcto servicio académico.		

Anexo N° 03. Modelo de Minería de Datos

ORGANIGRAMA.



1. Estudio de Viabilidad

El objetivo del estudio de viabilidad del proyecto es el análisis de un conjunto concreto de necesidades para proponer una solución a corto plazo, que tenga en cuenta restricciones económicas, técnicas, legales y operativas.

Viabilidad Técnica

Entre los recursos tenemos al personal que estará en el proyecto, así como los recursos de hardware y software y sus respectivas licencias.

Tabla 01: Recurso

TI

Requerimientos de Recursos	Mes 1	Mes 2	Mes 3	Mes 4
MS Office 2016	1	1	1	1
SQL Server 2008 R2 Enterprise Edition	1	1	1	1
QlikView	1	1	1	1
Computadora Personal	4	4	4	4
Servidor	2	2	2	2

Fuente: Elaboración Propia

Viabilidad Económica

Análisis de Costos

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

Costos de Inversión

Tabla 02: Costos de Personal

Recursos	Costo
Jefe de Proyecto	6,000.00
Experto en minería de datos	3,800.00
Experto en Inteligencia de Negocio	3,800.00
Analista de Procesos	3,500.00
Jefe de Pruebas	2,400.00

Fuente: Elaboración Propia

Tabla 03: Costo por Uso de Equipo Informático

Recurso TI	Total de hora	Costo Por hora	Costo
Computadora Personal	3,328	2.00	6,656.00

Fuente: Elaboración Propia

Tabla 04: Costos de Capacitación de Personal

Recurso	Cost/hora	Total	Costo
Capacitación de Personal	65.22	15	978.3

Fuente: Elaboración Propia

Tabla 05: Costos de Suministros e Insumos

Material	Cantidad	Costo Unitario	Subtotal
Papel Bond (millar)	2	30.0	60.0
Lapiceros	6	0.80	4.8
Corrector ortográfico	2	5.00	10.0
Engrapador	1	20.0	20.0
Perforador	1	20.0	20.0
Folder de manila	30	0.50	15.0
Clips (caja)	2	4.00	8.0
Cinta adhesiva	2	1.00	2.0
Tijera	2	0.50	1.0
TOTAL			140.80

Fuente: Elaboración Propia

Costos De Operación

Tabla 06: Costo por Uso de Servicios

Suministros	Costo
Energía Eléctrica	800.00
Conexión a Internet	1,080.00
Otros Gastos	500.00
TOTAL	2,380.00

Fuente: Elaboración Propia

Tabla 07: Otros Costos Variables

Suministros	Cantidad	Costo Unitario	Subtotal
Disco Duro Extraibles	1	250.00	250.00
Otros			100.00
TOTAL			350.00

Fuente: Elaboración Propia

Costos de Mantenimiento

Tabla14. Costos de Personal para Mantenimiento

Cargo	Días	Horas	Pago Por Horas (S/.)	Total (S/.)
Soporte técnico	10	80	10.0	800.00
Analista de Sistemas	10	80	10.0	800.00
Analista Programador de Sistemas	30	240	10.0	4,000.00
TOTAL				5600.00

Fuente: Elaboración Propia

Tabla 08: Costos Por Uso de Servicio

Suministros	Horas	Cantidad	Costo Por Horas	Total
Energía Eléctrica	80	3	0.5	120.00
Conexión a Internet	100	3	0.3	90.00
TOTAL				210.00

Fuente: Elaboración Propia

c) Análisis de Beneficios

Los proyectos de minería de datos, son siempre un reto para sus precursores, pues al realizar el análisis costo/beneficio respectivo, por un lado, los costos son fuertes en el primer año, mientras los beneficios

potenciales se presentan paulatinamente a mediano y largo plazo, esto es, no antes de seis meses.

Además, el jefe de proyecto se presenta ante la dificultad de expresar los beneficios concretos del proyecto en valores medibles. Veamos el siguiente ejemplo:

Cuando se presenta un sistema de facturación, es obligatorio exponer beneficios como aumento de la facturación en 40% por ejemplo. Cuando vemos una propuesta para inventarios, podemos exponer beneficios como disminución de robo; utilización de un método de valuación de inventarios que coloca el costo justo; aumento de los ingresos y egresos en un 50% por la automatización de órdenes de ingresos y salidas.

Pero cuando se habla de los beneficios de Minería de datos, los beneficios se circunscriben de los siguientes hechos:

- Aumento de la productividad: minería de datos es una solución orientada al usuario. Esto implica que el usuario tiene el poder de administrar su información: genera consultas y reportes por cuenta propia. Esta situación genera una descarga considerable de trabajo para el área lo cual permite al gerente utilizar este recurso disponible para otras tareas que demanden más análisis.
- Aumento de los Ingresos: La minería de datos permite al área crear una cultura de trabajo en equipo que permita el establecimiento de estrategias fundamentadas en los índices de mercadeo e indicadores confiables para el aumento de ingresos.
- Aumento de las Utilidades: Con la minería de datos, se puede centrar la atención en los clientes de mayor consumo y en los productos que tiene un margen de ganancia mucho mayor.
- Aumento de la Cartera de Clientes: Por medio de la minería de datos se puede definir el costo de un cliente nuevo.

d) Viabilidad Operativa

El desarrollo del estudio de factibilidad operativa comprende una probabilidad de que la nueva herramienta que se use se diseñe como se supone. Para tal hecho se debe considerar algunos aspectos de la factibilidad operacional como se mencionan a continuación:

Primero, un nuevo modelo de minería de datos puede ser demasiado complejo para los usuarios de la organización o los directivos de la empresa. Si llegase a serlo, los usuarios pueden ignorar el sistema o bien usarlo en tal forma que cause errores o fallas en el sistema.

Segundo, un modelo de minería de datos puede hacer que los usuarios se resistan a él como consecuencia de una técnica de trabajo, miedo a ser desplazados, u otras razones. Para cada una de estas alternativas se debe explorar con cuidado la posibilidad de resistirse al cambio o al nuevo uso del modelo de minería de datos.

Tercero, un nuevo modelo de minería de datos puede introducir cambios demasiado rápido para permitir al personal adaptarse a él y aceptarlo. Un cambio repentino que se ha anunciado, explicado y “vendido” a los usuarios (directivos) con anterioridad puede crear resistencia. Sin importar qué tan atractivo pueda ser un modelo de minería de datos en su aspecto económico si la factibilidad operacional indica que tal vez los usuarios no aceptarán el CMI o que su uso resultará en muchos errores o en una baja en la moral, el CMI tal vez no debe implantarse para evitar causar malestar en el desarrollo de las actividades de la empresa.

Una última consideración es la probabilidad de la obsolescencia tecnológica en la implantación del CMI. La tecnología que ha sido anunciada pero que aún no está disponible puede ser preferible a la tecnología que se encuentra en una o más de las alternativas que se están comparando. Cambios anticipados en las prácticas o políticas administrativas pueden hacer que un nuevo CMI sea obsoleto muy pronto. Desde el punto de vista operativo, que el impacto del nuevo CMI sobre la municipalidad en las cuales será positivo y sin grandes trabas debido a:

En primera instancia, la idea surge de la necesidad de tener información sobre los índices actuales o indicadores en base a los objetivos de la organización de la empresa; por lo cual, éste sistema se enfoca a resolver un problema concreto y que fija un punto de partida a la resolución del problema en el proceso de la toma de decisiones, la implementación del mismo no representa un cambio radical ya que no existe un sistema previo relativo al tema, de esta manera habiendo elaborado un detallado análisis de cada uno de los aspectos relacionados con la parte Técnica, Económica y Operacional del Estudio de Factibilidad, podemos concluir que el Proyecto es posible de desarrollar y teniendo en cuenta que los recursos que se requieren son de fácil adquisición, se puede determinar que el Proyecto es Factible en su Implementación.

e) Viabilidad Legal

Dentro del marco legal no se encuentra impedimentos en el desarrollo de un modelo de minería de datos siempre y cuando se cumpla con los derechos de autor, y las normas políticas de la organización de la empresa, de acuerdo a su reglamento interno el cual no conlleva a la protección de la información interna por ser de gran valor para la organización.

Si hacemos una revisión de los diferentes aspectos que compromete la formulación de estudios de pre inversión, encontramos en cada paso un cuerpo de legislación que determina la situación entre los diferentes actores de las relaciones comerciales, técnicas, financieras e institucionales que afectan la ejecución, operación y hasta la liquidación del proyecto

f) Vida útil del proyecto

El flujo de caja es un estado de cuenta básico que es construido para un numero específico de años, meses o de vida útil del proyecto. Esta vida útil depende de la capacidad del proyecto de generar una renta económica. Es muy importante que la vida útil de proyecto no se confunda con aquella de los activos adquiridos para el funcionamiento del mismo. En algunos

casos, la vida útil de los activos será menor a la del proyecto y, en otros, será mayor.

Ahora bien, como se mencionó anteriormente, un flujo de caja se debe construir hasta el momento en que las rentas económicas del proyecto sean iguales a cero. Sin embargo, este momento no es fácil de calcular a priori, por lo que se puede ser necesario utilizar una vida útil preestablecida dependiendo del giro del negocio y de sus características, tales como el sector donde se ubique, el tiempo que el inversionista desee permanecer en el negocio o algún otro indicador, una empresa constructora por ejemplo, asociará la vida útil de un proyecto al tiempo establecido en el contrato de construcción; un negocio de servicios, al periodo por el cual haya adquirido una concesión o por el tiempo por el que se obtiene una franquicia. También es posible atar la vida útil al tiempo de vida de un activo fijo, cuando la inversión del mismo es de una magnitud considerable. Así, en el caso de la instalación de un molino, la inversión en la maquinaria será de tal magnitud que el horizonte del proyecto dependerá totalmente de la vida útil del molino adquirido. Un caso similar es el de un agricultor que ha arrendado la tierra por un tiempo determinado, pasado el cual no sabe si podrá a volver a arrendarla.

Las compañías gastan mucho dinero en sistemas informáticos y, para obtener un beneficio de esa inversión, el software o el hardware debe utilizarse varios años. El tiempo de vida de los sistemas informáticos es muy variable, pero muchos sistemas grandes se pueden llegar a utilizar hasta más de 20 años. Muchos de estos sistemas antiguos aún son importantes para sus respectivos negocios, es decir, las empresas cuentan con los servicios suministrados por estos sistemas y cualquier fallo en estos servicios tendría un efecto serio en el funcionamiento de la organización. Entonces la vida útil de nuestro proyecto estaría dada por el avance de la tecnología y de las herramientas de inteligencia de negocio que van mejorando al transcurrir el tiempo. Es decir, nuestro proyecto tendrá una duración hasta que esta herramienta quede desplazada por una mejor.

2. Metodología

2.1 Metodología CRISP-DM.

La metodología CRISP-DM por sus siglas en inglés (*CRoss-Industry Standard Process for Data Mining*), es una metodología relativamente joven que nace no mucho tiempo después de que la Minería de Datos se forje como disciplina importante dentro del análisis de los Activos de Información que posee la Empresa.

Esta herramienta de trabajo surge de la necesidad de aprender nuevas técnicas para aplicar y comprender de mejor manera a la Minería de Datos y sus resultados basándose en un proceso jerárquico ya que se compone de diferentes niveles o tareas ofreciendo resultados a corto tiempo por lo que según algunos textos que tratan sobre Business Intelligence lo consideran como una metodología de Minería de Datos para inexpertos. Su construcción está basada en experiencias prácticas más que de teoría, por ser una metodología abierta se la puede usar con varias herramientas de Business Intelligence. Se caracteriza por comenzar su análisis desde una perspectiva global enfatizando en el conocimiento del negocio, de esta forma la metodología está más apegada al proyecto que se realiza en la Empresa como primera fase.

2.2 Fases del Modelo de Referencia.

i. Comprensión del Negocio.

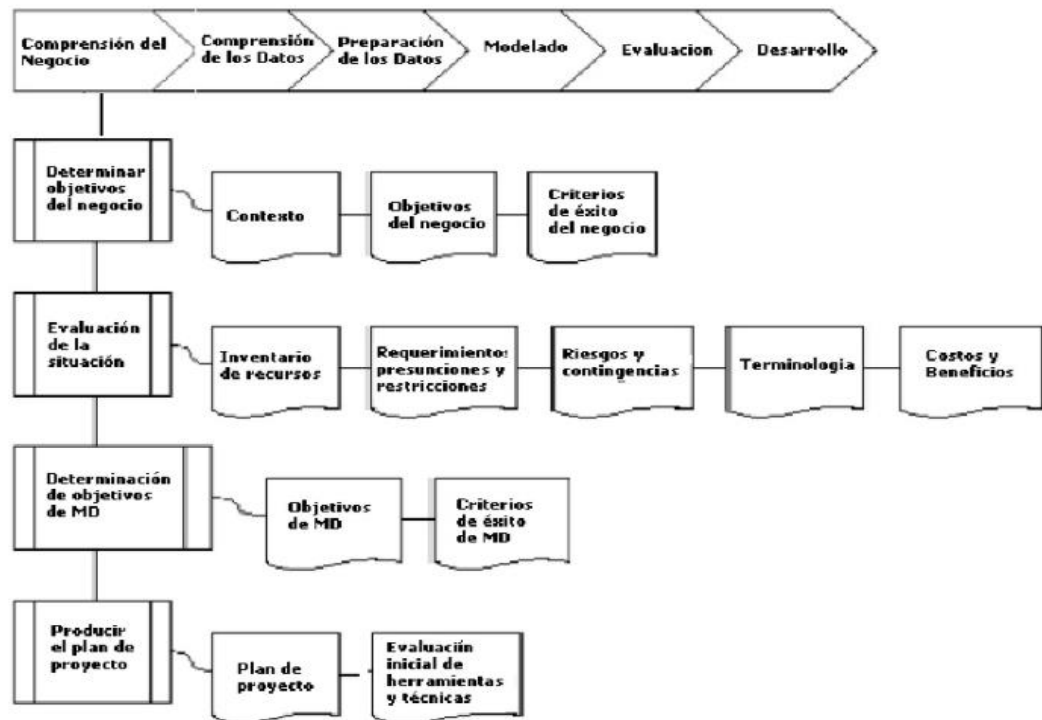
Corresponde a la primera fase del problema, el entendimiento del negocio y sus objetivos es de vital importancia para comenzar con un proyecto de Minería de Datos ya que se debe tener una visión de tipo empresarial y competitiva conociendo el entorno de la Empresa. Se trata de comprender las necesidades que tiene el cliente y de que forma la Empresa puede aportar para satisfacer las mismas; en este caso se pretende ofrecer servicios que la Empresa dispone y que al cliente le interese adquirir. Es aquí donde se necesita del conocimiento del entorno de trabajo, de esta manera se beneficiará la Empresa directamente y obliga a seguir innovando en proyectos afines. En este punto se recomienda establecer un

grupo de trabajo donde se pueda crear una lluvia de ideas tanto de las personas que ya conocen el negocio como de las que aún no lo conocen y pueden dar una visión desde la perspectiva del cliente. Lo que se busca es establecer nuevos propósitos para poder captar mayor cantidad de clientes o lanzar nuevos servicios o actualizar según los avances tecnológicos los ya existentes. Para esto se debe identificar las necesidades de los clientes además de sus características y la razón por la que prefirieron a esta Empresa. Una manera efectiva de incrementar las ventas de los servicios que ofrece la Empresa. es por medio de la creación de estrategias de marketing que ofrezcan planes de servicios que le interesan al cliente, estos planes pueden incluir paquetes de servicios que estén relacionados entre ellos, de esta forma el cliente se interesa por el paquete que se le está ofertando generando un mayor interés en el consumidor por adquirir el servicio y aumentando las ganancias para la Empresa. Con estrategias como la esta se pretende afianzar la relación entre el cliente y la Empresa dando a conocer e interés que tiene la compañía por mantener satisfechos a sus clientes fortaleciendo el concepto de lealtad hacia la Empresa. Un aspecto importante a tomar en cuenta es que un solo cliente puede estar interesado en varios servicios de los que ofrece la Empresa ya sea en distintos instantes de tiempo o no; al tratarse de un solo cliente consumiendo varios servicios hace que las estrategias de venta se concentren en el mantenimiento de cliente y estimularlo a que adquiera más servicios o incluso a colaborar con el incremento de clientes; esto se puede producir al momento en que el cliente se encuentra satisfechos y comienza a promocionar a la Empresa y sus servicios casi de forma inconsciente con sus allegados. La Empresa debe contar con personas autorizadas y capacitadas para decidir qué estrategia de marketing emplear para la promoción de sus servicios.

Esta fase se divide en:

- Determinar objetivos del negocio.
- Evaluación de la situación.
- Determinación de objetivos de Minería de Datos.

- Producir el plan de proyecto.



ii. Determinar objetivos del Negocio.

Para este aspecto es importante que se tenga un punto de vista Empresarial de tal manera que se enfoque en los objetivos competitivos que tiene la Empresa, comprende también el claro entendimiento del problema de la Empresa y lo que se busca obtener. Según la metodología, el hecho de ubicar el contexto dentro de la determinación de los objetivos de la Empresa, implica que se debe ubicar al inicio del proyecto la información de la Empresa a la cual se está aplicando la metodología. Información referente a la actividad a la que se dedica la Empresa, los productos y/o servicios que ofrece la Empresa y sobretodo tener en cuenta las necesidades de los clientes y las razones por la cual podría interesarle los servicios de la competencia.

iii. Evaluación de la Situación.

Se ubican todos los Sistemas de Información que son de propiedad de la Empresa para tener en cuenta los recursos humanos e informáticos con los que se puede contar para realizar el proyecto. Parte de la evaluación

comprende la identificación de restricciones y mantenerse al tanto de la información a la que se tiene acceso y a la información a la que no, los permisos que se otorgarán al usuario, etc.

Determinación de Objetivos de Minería de Datos

Comprende una definición de objetivos usando términos más técnicos aplicados ya a definiciones apegadas a la Minería de Datos, esa es la diferencia con el objetivo del negocio.

Producir el plan de Proyecto.

Para llegar a esta etapa se deberá cumplir con los objetivos de Minería de Datos, este a su vez servirá para cumplir los objetivos de negocio, además de señalar cuales son las herramientas que se usarán, estableciendo si dicha herramienta ya posee la Empresa o requieren ser adquiridas.

Comprensión de los Datos.

La comprensión de datos hace referencia a una recopilación inicial de los datos de la Empresa que luego serán convertidos en información de esta manera poco a poco se produce la familiarización de los datos. Este proceso pretende dar a conocer alguna anomalía dentro de los sistemas de Información como inconvenientes en la calidad de los datos. Se puede considerar una reducción del volumen de datos, descartando información que para el presente objetivo no puede ser de mucha importancia. Identificar variables importantes, evitando la redundancia de datos sin que este afecte al contenido mismo de la información. Dentro de la Empresa se aplicará la comprensión de datos directamente con el sistema que se encuentra en AS-400 que es donde esta gran parte de la información de la Empresa, y el estado en el que se encuentra la misma, desde aquí se pueden hacer consultas (query) y comenzar a descartar información y maximizar el entendimiento. Se debe tomar en cuenta los servicios que presta la Empresa e ir asociando la información que se indague con alguno de los servicios.

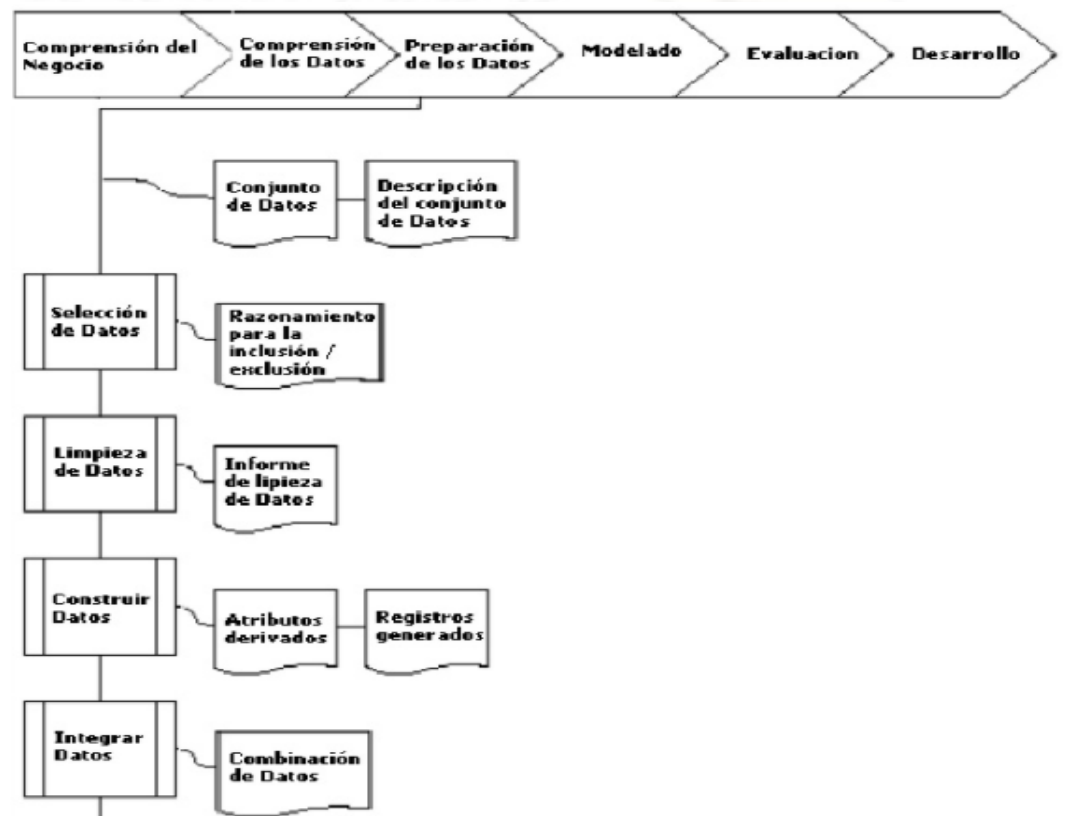
iv. Preparación de los Datos.

Esta fase toma más tiempo que las anteriores, más de la mitad de todo el trabajo y requiere de mayor esfuerzo, esta fase incluye todas las acciones que se usarán para los diferentes procesos consulta de la información, además se integran etapas importantes como la Transformación y Limpieza de Datos. Se han seleccionado ya los datos con los que se trabajará durante todo el proceso del proyecto y cubre a todas las actividades que se realicen para cumplir con el proyecto. Para el perfeccionamiento de los datos con los que se va a trabajar se tomará en cuenta sobre ETL aplicando algoritmos basados en Inteligencia Artificial como una de las bases de la Minería de Datos seleccionados para sondear información y obtener resultados favorables, ya que si tomamos los datos tal cual están almacenados en la Base de Datos no entregarán resultados favorables. El verdadero valor de la información es el que se obtiene luego de un proceso aplicando una técnica que de soporte para la posterior toma de decisiones. Para la selección de los algoritmos debemos tener muy claro el objetivo que se está buscando, el volumen de información con el que se va a necesitar, finalmente realizar reportes de la información obtenida, debiendo establecer un margen de error admisibles.

En este caso se trabajará con la mayoría de variables de tipo numérico, desde donde se partirá para el estudio y aplicación de los diferentes algoritmos que se explican más adelante. Usando las variables definidas se buscará establecer patrones de comportamiento de los clientes de los diferentes servicios ya mencionados, de tal forma mantenerlos plenamente identificados tanto a los clientes potenciales como a los que están por abandonar la institución, sería ideal estar al tanto de los motivos por los cuales algunos clientes no están consumiendo el servicio y mejor aún; saber por qué las personas prefieren otras instituciones que ofrecen servicios similares. Todo este trabajo se debe realizar sin dejar de lado el ruido o error que trae consigo un proyecto de Minería de Datos, es precisamente por eso que se debe tener la capacidad de identificar posible

información errónea que la herramienta pueda arrojar como resultado. Esta fase se divide en:

- Selección de los datos.
- Limpieza de los datos.
- Construir de los datos.
- Integrar de los datos.
- Formatear los datos.



El conjunto de datos representa el grupo de datos con los que se va a trabajar en el proyecto, este debe ser descrito a detalle especificando su intervención dentro del proyecto.

Selección de los Datos.

Consiste en la designación de datos a utilizarse durante el desarrollo del proyecto, al mencionar selección de datos se está haciendo referencia a los esquemas, tablas, registros además diferenciar información a la que se tiene acceso y cuáles son los permisos que de los que se dispone. La selección de datos incluye también los datos que serán incluidos o excluidos en el proyecto y los motivos para ello.

Depende mucho de los objetivos que se hayan planteado para el proyecto además de que en el transcurso del desarrollo se encuentran nuevos datos que resultan ser significativos.

Limpieza de los Datos.

Evitar que la información llegue distorsionada o duplicada así el proyecto perderá veracidad, se debe contar con datos que contengan un nivel de calidad aceptable para poder aplicar las diferentes técnicas seleccionadas. Dentro del informe de limpieza de datos se establece las acciones que han sido tomadas para mejorar la calidad de la información cual ha sido el impacto que ha tenido dentro del proyecto.

Construir de los Datos.

Incluye operaciones para preparación de los datos como la modificación de datos o añadir registros si fuese necesario. Ha estos cambios se denominan atributos derivados ya que su origen proviene de tablas o registros ya existentes, en el caso de que se hayan creado tablas o registros nuevos se deberá justificar su creación, describir cuáles son sus atributos y el objetivo de su creación.

Integrar de los Datos.

Técnica mediante la cual se combina la información de las diferentes tablas de las que se disponga y que pueden estar relacionadas entre si pero que contiene información diferente mediante algún atributo o cruzando diferentes tablas, de esta forma se produce la combinación de tablas.

v. Modelado.

Dentro de esta fase se seleccionan varias técnicas de modelado luego estudiar su funcionamiento, por lo que la nueva técnica se agregará a la información que se ha obtenido en el proceso anterior con el objetivo de resolver un problema de Minería de Datos. Existen varias formas de resolver un mismo problema de MD pero se debe acertar a la técnica que más se apegue a las necesidades de la consulta. Es muy frecuente observar que cuando el proyecto esté en esta fase se tenga que regresar al ciclo de "Preparación de Datos" por motivos de formatos de datos, en

este caso se procede a estudiar nuevamente el porqué de la necesidad de retroceder además de proveer de soluciones para tratar de evitarlo. El proceso de Modelado muestra ya los patrones conductuales como resultado de la aplicación de las técnicas, cabe señalar que la Base de Datos de la Empresa se actualiza diariamente es por eso que los datos históricos son parte importante dentro de este proyecto ya que se osaran datos que correspondan a fechas específicas. Se escoge las técnicas de modelamiento en base a los siguientes criterios. Ser apropiada al problema. Disponer de Datos adecuados. Cumplir los requerimientos del problema. Tiempo necesario para obtener un problema. Conocimiento de la técnica. Esta fase se divide en:

- Seleccionar la técnica de modelamiento.
- Generación de pruebas de modelo.
- Construcción del modelo.
- Evaluación del modelo.

Seleccionar la Técnica de Modelamiento.

Seleccionar la técnica que se aplicará, se trata de una técnica específica que se tendrá que conocer su funcionamiento e interpretar los resultados. En el caso de que se aplique más de una técnica se tendrá que especificar por separado documentando cada una delas técnicas usadas para una mejor comprensión.

Generación de Pruebas de Modelo.

Se debe formar un medio que describa el procedimiento para verificar la validez de las técnicas que están siendo usadas. La metodología indica que se debe tener por separado datos para que sean para hacer pruebas. Generalmente suelen existir herramientas de software que no soportan grandes cantidades de datos es por eso que es óptimo seleccionar cierta muestra para la aplicación de pruebas, la ventaja de trabajar con datos de pruebas es la velocidad de ejecución de respuesta, así se puede manipular el software que se esté usando con un grado mayor de confianza.

Construcción del Modelo.

Una vez escogidos los datos se debe aplicar el modelo seleccionado ajustándolos parámetros necesarios y describir los modelos obtenidos añadiendo la interpretación que se tenga de cada uno de ellos.

Evaluación del Modelo.

Esta tarea le corresponde a un experto puede tratarse de la persona quien esté a cargo del liderazgo del proyecto o el mismo desarrollador, el cual debe decidir si las técnicas usadas han sido las más idóneas y si los resultados obtenidos son viables o no para cumplir con los objetivos propuestos, esto se hará conjuntamente con personal que conozca a fondo el negocio. Incluso en esta fase se habla de un Ing. en Minería de Datos como experto que puede también dar opiniones sobre el negocio desde una perspectiva más técnica si la Empresa lo tuviese, pero no es el caso.

vi. Evaluación.

Para esta fase los modelos que fueron escogidos deben estar ya implementados dentro de la Empresa, cabe señalar que se debe hacer un recuento de todos los pasos que se siguieron hasta llegar a la evaluación con el fin de asegurarse de que no se ha pasado por alto algún detalle. En el momento en que el proyecto se encuentra en esta fase, se entiende que los objetivos del proyecto fueron alcanzados desde el punto de vista Empresarial antes el punto de vista técnico, se puede hablar de un éxito o fracaso en el progreso de toda la metodología. Se recomienda documentar todo el proceso ya que dentro de la fase de evaluación se pudo cometer algún error que pueda provocar el regreso a las fases anteriores, esto puede pasar durante el desarrollo de un proyecto de Minería de Datos. Esta fase se divide en:

- Evaluación de los resultados.
- Proceso de revisión.
- Determinación de los próximos pasos

Evaluación de los Resultados.

Se trata de valorar, como se mencionó anteriormente si las técnicas aplicadas y los diferentes modelos dieron como resultado la información que se necesitaba para cumplirlas metas del proyecto, al mismo tiempo hacer una evaluación de los recursos usados y el tiempo invertido. Los resultados se consideran como satisfactorios siempre y cuando tengan un margen de aceptación razonable para la Empresa, caso contrario se debe encontrar los parámetros idóneos para una óptima deducción.

Proceso de Revisión.

Los modelos ya han sido aplicados y los resultados ya están catalogados como satisfactorios o no, en ambos casos se recomienda plantearse interrogantes como; ¿Se usaron los datos necesarios? ¿Olvidó plantear alguna otra variable que pueda representar cambios dentro de los resultados? El proceso de revisión debe realizarse en presencia de todo el equipo que forma parte del proyecto, de esta manera se puede tomar en cuenta todos los puntos de vista y se consideran las falencias o aciertos de los resultados. Generalmente en el proceso de revisión surgen cambios, este hecho debe considerarse totalmente normal ya que cada vez surgen más necesidades e ideas siempre y cuando estén dentro de lo permitido y cuente con el apoyo y justificación de todo el grupo de trabajo

Determinación de los Próximos Pasos.

Para la determinación de las siguientes acciones a tomar se cuenta con las personas que aportaron para el desarrollo del proyecto, dicho grupo deberá tomar en cuenta las acciones futuras y los posibles resultados que se espera obtener. En este caso se plantea ejecutar cada que sea necesario los paquetes desarrollados así los paquetes están diseñados para generar información según la fecha deseada como se indica. De esta manera se consigue una reutilización de los paquetes anteriormente realizados.

vii. Despliegue o Desarrollo.

El despliegue del proyecto no significa que el propósito haya terminado, ya que se deben hacer pruebas de su implementación, así como un

seguimiento periódico. Los resultados que presenta deben ser claros de tal manera que se puedan comprender fácilmente por personas que no tengan mucha intervención en el área de la Informática, sino también de profesionales o no que pertenezcan a otras áreas. Tomando en cuenta lo anterior, lo más óptimo es que se presenten resultados de una manera visual usando cuadros estadísticos que representen las variables usadas y su aporte para la Empresa. La complejidad de la fase de despliegue dependerá de los requisitos que han propuesto así puede ir desde la simple presentación de un informe hasta la demostración del sistema ya implementado. Se busca dejar camino abierto para un posterior análisis, dejar de base este plan para seguir adelante con la obtención de nuevo conocimiento que sea de verdadera utilidad para la Empresa

Esta fase se divide en:

- Desarrollo del plan
- Plan de supervisión y mantenimiento.
- Producción del reporte final
- Revisión del proyecto.

Desarrollo del Plan.

Consiste en desplegar una técnica que permita poner en pie el trabajo desarrollado y elaborar un informe detallado incluyendo cronograma de actividades y cuáles son las medidas que se han de tomar.

Plan de Supervisión y Mantenimiento.

Trata de la preparación de una táctica que permita un control sobre los resultados que se va obteniendo durante el progreso además de ciertas tutorías con las personas encargadas de los sistemas informáticos sobre la nueva información que generalmente aparece dentro de un proyecto de Minería de Datos.

Producción del Reporte Final.

El informe final del proyecto se basa en un resumen de los pasos que se han implementado a lo largo del proyecto, así como los inconvenientes encontrados. El reporte final debe ser realizado por el líder del proyecto

conjuntamente con su equipo de trabajo. El formato de elaboración del reporte final corresponde a las políticas de la Empresa donde se esté aplicando el proyecto.

Revisión del Proyecto.

Revisar el proyecto e identificar lo que esta correcto y lo que no lo está y debe ser modificado si las circunstancias lo permiten, estas pueden ser: Autorización por parte de los directores. Tiempo. Recursos Humanos y Económicos.

Entre los productos utilizados están:

Pentaho

Pentaho se define a sí mismo como una plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos tal como ha sido concebido desde el principio.

Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos, actividades, además es capaz de presentar y entregar la información adecuada en el momento adecuado.

Su modelo de ingresos parece estar orientado a los servicios (soporte, formación, consultoría y soporte a ISVs y distribuciones OEM) aunque en alguno de los documentos y páginas que hemos examinado se mencionan algunas funcionalidades “Premium” que hacen pensar en ingresos por futuras versiones o funcionalidades de pago.

En su web presenta una organización por productos: Reporting, Analysis, Dabsboards y Data Mining, acompañado por dos introducciones: a la plataforma y a los productos. En dichas introducciones se hace mención específica al workflow como una de las capacidades BI claves de la plataforma.

Pentaho Analysis Services: cuyo nombre código Modrian es un servidor

OLAP (procesamiento analítico en línea) escrito en Java. Es compatible con el MDX (expresiones multidimensionales) y el lenguaje de consulta XML para el Análisis y especificaciones de la interfaz olap4j.

Pentaho Reporting: Consiste en un motor de presentación, capaz de generar informes programáticos sobre la base de un archivo de definición XML. Sobre esta solución se han desarrollado muchas herramientas, por ejemplo informes, diseñadores de interfaz gráfica de usuario, y asistentes tipo wizard. Un uso notable de esta herramienta es el Generador de informes para OpenOffice.org

Pentaho Data Mining: Es una envoltura alrededor del proyecto Weka. Es una suite de software que usa estrategias de aprendizaje de máquina, aprendizaje automático y minería de datos. Cuenta con series de clasificación, de regresión, de reglas de asociación, y de algoritmos de clustering, para así apoyar las tareas de análisis predictivo.

Pentaho Dashboard: Es una plataforma integrada para proporcionar información sobre sus datos, donde se pueden ver informes, gráficos interactivos y los cubos creados con las herramientas Pentaho Report Designer.

- 6.2.1 **Pentaho para Apache Hadoop:** Es un conector de bajo nivel para facilitar el acceso a MUY grandes volúmenes manejados en el proyecto Apache Hadoop, la Suite de Pentaho BI para Hadoop permite abordar los mayores desafíos que experimentan los usuarios de Hadoop -, sobre su empinada curva de aprendizaje técnico, la falta de personal técnico cualificado y la falta de disponibilidad de las aplicaciones de desarrollo y despliegue para llevar a cabo la integración de datos e inteligencia de negocios con Hadoop.+



SQL Server Business Intelligence

SQL (por sus siglas en inglés **Structured Query Language**; en español **lenguaje de consulta estructurada**) es un lenguaje específico del dominio que da acceso a un sistema de gestión de bases de datos relacionales que permite especificar diversos tipos de operaciones en ellos. Una de sus características es el manejo del álgebra y el cálculo relacional que permiten efectuar consultas con el fin de recuperar, de forma sencilla, información de bases de datos, así como hacer cambios en ellas.

Originalmente basado en el álgebra relacional y en el cálculo relacional, SQL consiste en un lenguaje de definición de datos, un lenguaje de manipulación de datos y un lenguaje de control de datos. El alcance de SQL incluye la inserción de datos, consultas, actualizaciones y borrado, la creación y modificación de esquemas y el control de acceso a los datos. También el SQL a veces se describe como un lenguaje declarativo, también incluye elementos procesales.

SQL fue uno de los primeros lenguajes comerciales para el modelo relacional de Edgar Frank Codd como se describió en su papel de 1970 *El modelo relacional de datos para grandes bancos de datos compartidos*. A pesar de no adherirse totalmente al modelo relacional descrito por Codd, pasó a ser el lenguaje de base de datos más usado.

SQL pasó a ser el estándar del Instituto Nacional Estadounidense de Estándares (ANSI) en 1986 y de la Organización Internacional de Normalización (ISO) en 1987. Desde entonces, el estándar ha sido revisado para incluir más características. A pesar de la existencia de ambos estándares, la mayoría de los códigos SQL no son completamente portables entre sistemas de bases de datos diferentes sin ajustes.

SQL es un lenguaje de acceso a bases de datos que explota la flexibilidad y potencia de los sistemas relacionales y permite así gran variedad de operaciones.

Es un lenguaje declarativo de "alto nivel" o "de no procedimiento" que, gracias a su fuerte base teórica y su orientación al manejo de conjuntos de registros - y no a registros individuales - permite una alta productividad en codificación y

la orientación a objetos. De esta forma, una sola sentencia puede equivaler a uno o más programas que se utilizarían en un lenguaje de bajo nivel orientado a registros. SQL también tiene las siguientes características:

- **Lenguaje de definición de datos:** El LDD de SQL proporciona comandos para la definición de esquemas de relación, borrado de relaciones y modificaciones de los esquemas de relación.
- **Lenguaje interactivo de manipulación de datos:** El LMD de SQL incluye lenguajes de consultas basado tanto en álgebra relacional como en cálculo relacional de tuplas.
- **Integridad:** El LDD de SQL incluye comandos para especificar las restricciones de integridad que deben cumplir los datos almacenados en la base de datos.
- **Definición de vistas:** El LDD incluye comandos para definir las vistas.
- **Control de transacciones:** SQL tiene comandos para especificar el comienzo y el final de una transacción.
- **SQL incorporado y dinámico:** Esto quiere decir que se pueden incorporar instrucciones de SQL en lenguajes de programación como: C++, C, Java, PHP, Cobol, Pascal y Fortran.
- **Autorización:** El LDD incluye comandos para especificar los derechos de acceso a las relaciones y a las vistas.

Tipos de datos

Algunos de los tipos de datos básicos de SQL son:

- **Varchar:** Recibe cadena de palabras compuestas de letras, números y caracteres especiales.
- **Date:** una fecha de calendario que contiene el año (de cuatro cifras), el mes y el día.
- **Time:** La hora del día en horas minutos segundos (el valor predeterminado es 0).
- **Datetime:** la combinación de Date y Time.

Optimización

Como ya se dijo antes, y suele ser común en los lenguajes de acceso a bases de datos de alto nivel, SQL es un lenguaje declarativo. O sea, que especifica

qué es lo que se quiere y no cómo conseguirlo, por lo que una sentencia no establece explícitamente un orden de ejecución.

El orden de ejecución interno de una sentencia puede afectar seriamente a la eficiencia del SGBD, por lo que se hace necesario que éste lleve a cabo una optimización antes de su ejecución. Muchas veces, el uso de índices acelera una instrucción de consulta, pero ralentiza la actualización de los datos. Dependiendo del uso de la aplicación, se priorizará el acceso indexado o una rápida actualización de la información. La optimización difiere sensiblemente en cada motor de base de datos y depende de muchos factores.

Existe una ampliación de SQL conocida como FSQL (Fuzzy SQL, SQL difuso) que permite el acceso a bases de datos difusas, usando la lógica difusa. Este lenguaje ha sido implementado a nivel experimental y está evolucionando rápidamente.



Crystal Reports

Crystal Reports ha formado parte de Visual Studio desde 1993, y ahora es el estándar de elaboración de informes de Visual Studio. Se incluye en todas las copias de Visual Studio Professional y se integra directamente en el entorno de desarrollo.

Crystal Reports para Visual Studio incorpora la posibilidad de crear contenido interactivo con calidad de presentación al entorno de Windows. Con Crystal Reports para Visual Studio, puede crear informes complejos y profesionales en un programa basado en GUI. Después puede conectar el informe a casi todos los orígenes de base de datos, así como a datos proxy, como un conjunto de resultados (por ejemplo, un ADO.NET DataSet). Los asistentes del

diseñador de GUI le permiten establecer fácilmente los criterios de formato, agrupamiento y gráficos, etc.

Puede almacenar el informe en una aplicación Web o para Windows, con uno de los controles de visores de Crystal Reports para Visual Studio. La presentación de informes, tanto en clientes Windows como en HTML 3.2 ó 4.0, es muy interactiva y proporciona funciones como la profundización en gráficos, la exploración de informes y la búsqueda de texto.

Crystal Reports para Visual Studio incluye un SDK extenso. Puede utilizarlo para interactuar con el informe mediante programación en tiempo de ejecución, usando uno de los cuatro modelos de objetos posibles:

- CrystalReportViewer, el modelo de objetos más sencillo.
- ReportDocument, el modelo de objetos más completo.
- ReportClientDocument, el modelo de objetos más completo. Este modelo de objetos está disponible con Crystal Reports 2008 o con un servidor RAS.
- InfoObject, un modelo de objetos muy eficaz para la programación y configuración de informes en el marco de Crystal Reports Server o Business Objects Enterprise.

Los informes se pueden relacionar con el proyecto de Visual Studio de muchas formas:

- Incruste los informes directamente en el proyecto.
- Obtenga acceso a ellos externamente, desde un directorio de archivos.
- Obtenga acceso a ellos como servicio Web de informes desde un servidor remoto.
- Conéctelos como Crystal Services.
- Mediante la actualización a Crystal Reports Server puede tener acceso a ellos a través del servidor RAS.
- Mediante la actualización a BusinessObjects Enterprise, puede tener acceso a ellos a través de Page Server o del servidor RAS incluidos en el marco de BusinessObjects Enterprise.

A continuación, se describe un ejemplo de cómo deben funcionar los informes en un proyecto de Visual Studio:

Crystal Reports se utiliza para diseñar un informe que permita a los usuarios profundizar en un gráfico y filtrar información, en función de las necesidades. Ese informe se ha incluido en un proyecto ASP.NET y después se ha mostrado en un formulario Web con un control CrystalReportViewer que está enlazado a ese informe. Se puede interactuar con el informe mediante el uso del modelo de objetos ReportDocument.



Base de Datos del Colegio

El diseño de una base de datos consiste en definir la estructura de los datos que debe tener un sistema de información determinado. Para ello se suelen seguir por regla general unas fases en el proceso de diseño, definiendo para ello el modelo conceptual, el lógico y el físico.

- En el diseño conceptual se hace una descripción de alto nivel de la estructura de la base de datos, independientemente del SGBD (Sistema Gestor de Bases de Datos) que se vaya a utilizar para manipularla. Su objetivo es describir el contenido de información de la base de datos y no las estructuras de almacenamiento que se necesitarán para manejar dicha información.
- El diseño lógico parte del resultado del diseño conceptual y da como resultado una descripción de la estructura de la base de datos en términos de las estructuras de datos que puede procesar un tipo de SGBD. El diseño lógico depende del tipo de SGBD que se vaya a utilizar, se adapta a la tecnología que se debe emplear, pero no depende del producto concreto. En el caso de bases de datos convencionales relacionales (basadas en SQL para entendernos), el diseño lógico consiste en definir las tablas que existirán, las relaciones entre ellas, normalizarlas, etc...

- El diseño físico parte del lógico y da como resultado una descripción de la implementación de una base de datos en memoria secundaria: las estructuras de almacenamiento y los métodos utilizados para tener un acceso eficiente a los datos. Aquí el objetivo es conseguir una mayor eficiencia, y se tienen en cuenta aspectos concretos del SGBD sobre el que se vaya a implementar. Por regla general esto es transparente para el usuario, aunque conocer cómo se implementa ayuda a optimizar el rendimiento y la escalabilidad del sistema.

Código	Apellidos y Nom...	Sexo	Profesión	Experiencia	Hrs#_Capacita...	Hrs#_Tardanza	Grado_Académico	Situación
10000001	CURO RAMOS, ...	M	EP	5	850	2	MAGISTER	Contratado
2098787	DAMIAN ARTEA...	F	EP	4	550	3	LICENCIADO	Contratado
201345	GUERRA GUERR...	M	EP	2	500	2	MAGISTER	Contratado
456723	SOTO PALACIO...	F	EP	2	250	2	LICENCIADO	Contratado
12567	HUAMAN PÍÑAS...	F	ES	2	240	2	LICENCIADO	Contratado
2456678	PINEDA PÍÑAS...	F	EP	1	180	1	LICENCIADO	Contratado
678923	VILCAS ARTES...	F	EP	1	180	1	LICENCIADO	Contratado
453267	BOZA CERRON...	M	ES	1	172	1	LICENCIADO	Contratado
4887	CORDOVA PERE...	F	ES	2	200	2	LICENCIADO	Contratado
12398	CORDOVA HINO...	F	ES	3	350	3	MAGISTER	Contratado
9835	BELTRAN CERR...	M	ES	3	340	2	MAGISTER	Contratado
87235	HUAROC LANDE...	F	EP	3	350	2	MAGISTER	Contratado
2378653	MENDEZ LOPEZ...	M	EP	4	480	2	MAGISTER	Contratado
2349876	MENDEZ ESPINO...	M	ES	4	480	2	LICENCIADO	Contratado
12987	PARLONA JULCA...	M	ES	4	490	3	LICENCIADO	Contratado
34567	PERALTA VENTU...	M	ES	5	600	3	MAGISTER	Contratado
234989	SOTOL VILLAR...	M	ES	5	590	3	MAGISTER	Contratado
235678	ALANZA TORREJ...	M	EP	6	620	4	LICENCIADO	Contratado
13498	ROBLES ROBLES...	F	EP	7	680	4	MAGISTER	Contratado
234	VILLAR MONTES...	M	EP	8	820	2	MAGISTER	Contratado
236789	CHOCCA MONT...	M	EP	7	740	1	MAGISTER	Contratado
345671	TORRES TORRE...	F	ES	6	680	3	MAGISTER	Contratado
334578	POMA POMA, AL...	F	ES	5	600	4	MAGISTER	Contratado
2345	HUAMAN PEÑA...	M	EP	4	720	2	MAGISTER	Contratado
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

BDD Transaccional

Administrador de conexiones

Proveedor: OLE DB nativo\SQL Server Native Client 10.0

Nombre del servidor: localhost [Actualizar]

Conexión con el servidor

☒ Usar autenticación de Windows

☐ Usar autenticación de SQL Server

Nombre de usuario: []

Contraseña: [] ☐ Guardar mi contraseña

Establecer conexión con una base de datos

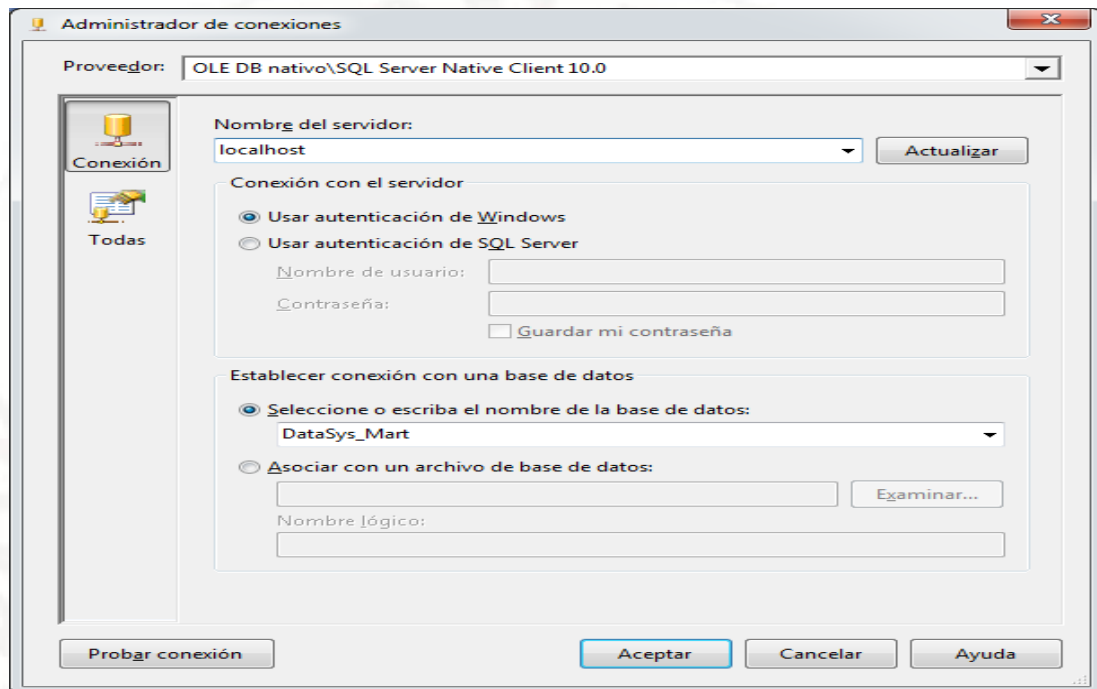
☒ Seleccione o escriba el nombre de la base de datos: DataSys

☐ Asociar con un archivo de base de datos: [] [Examinar...]

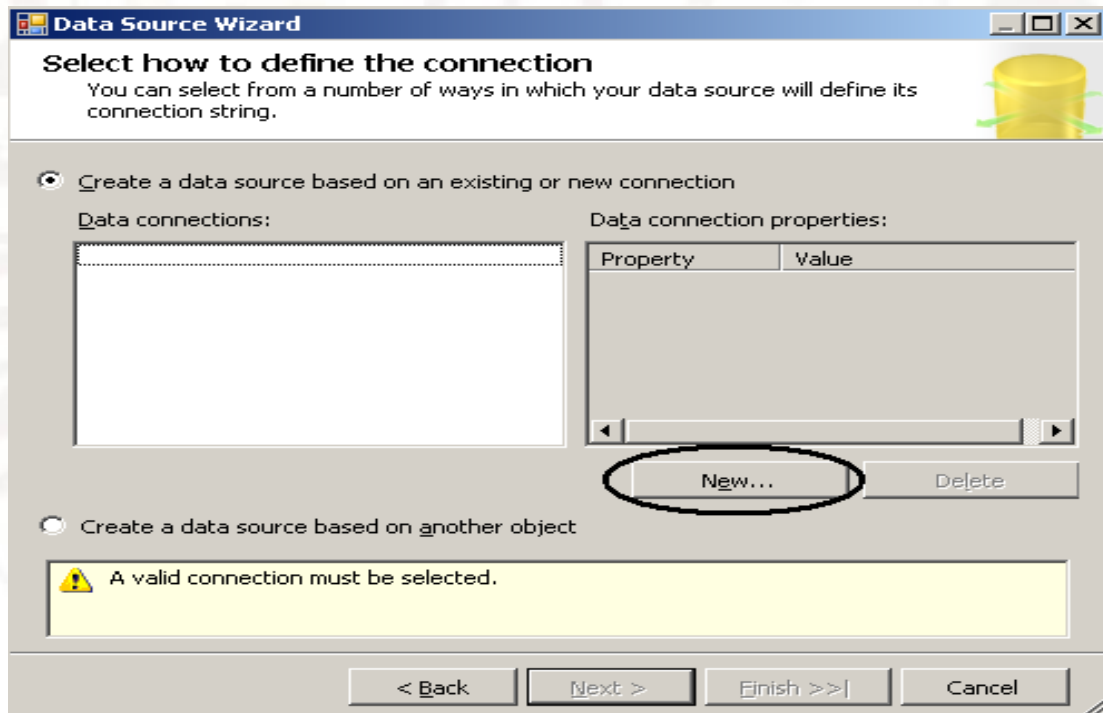
Nombre lógico: []

[Probar conexión] [Aceptar] [Cancelar] [Ayuda]

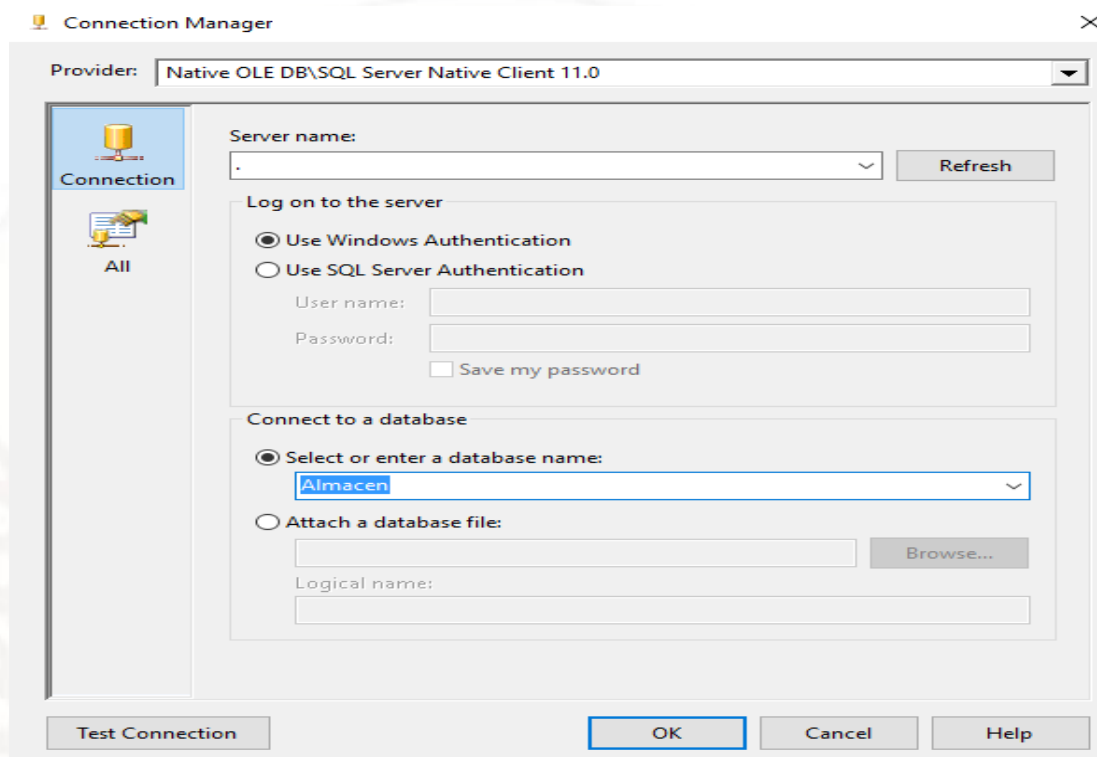
BDD Multidimensional



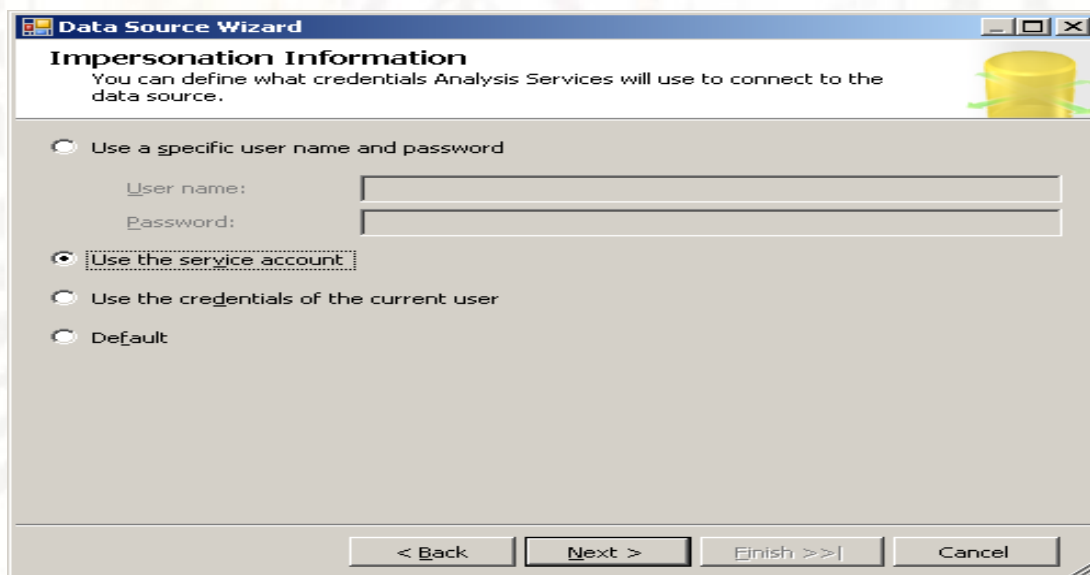
Origen de Datos



Creando Conexión

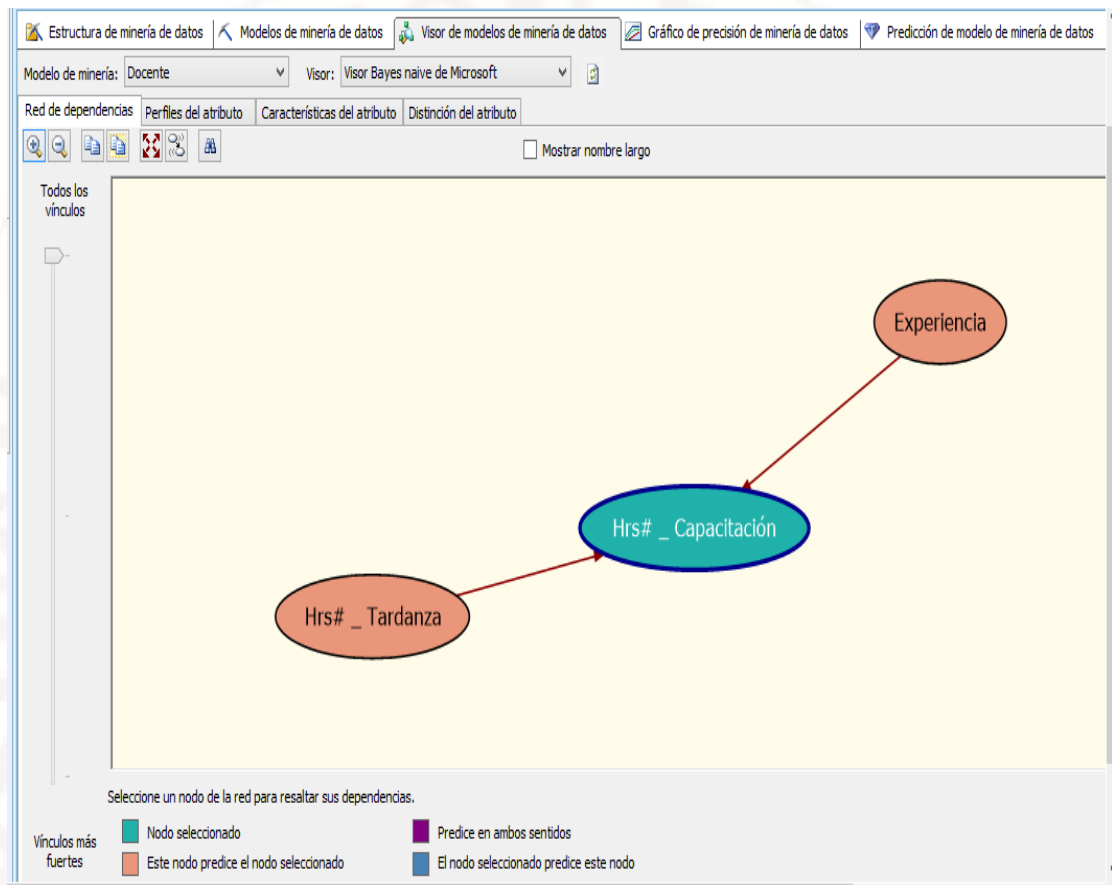


Administrando Cuenta de Conexión

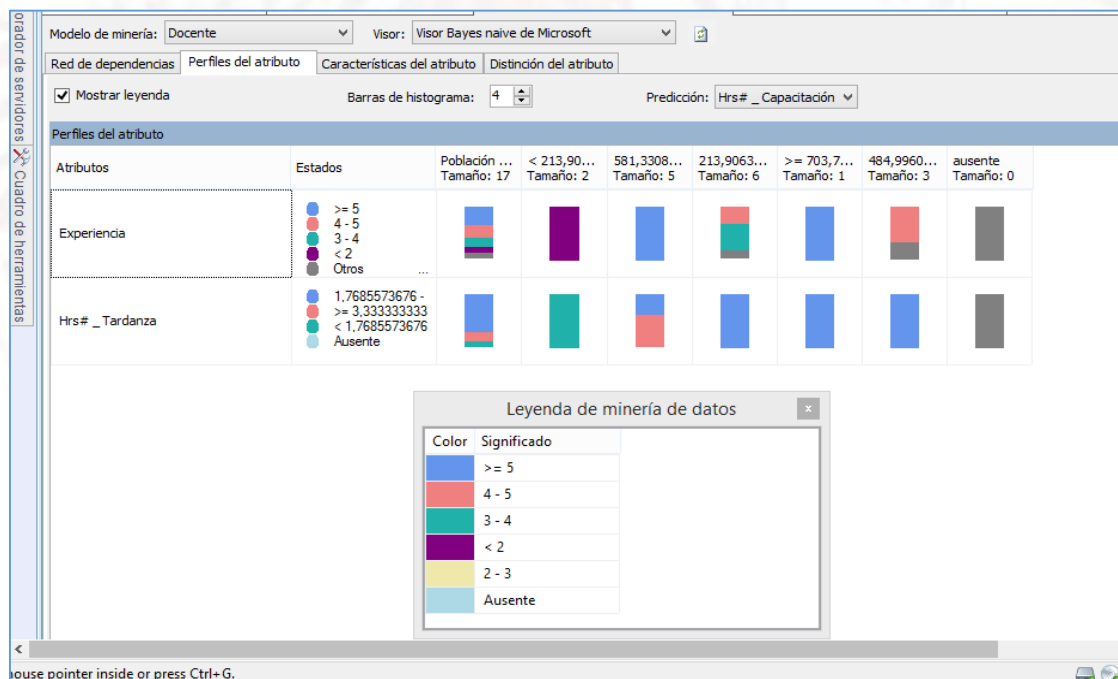


3. MINERIA DE DATOS DOCENTE: VISOR BAYES NAIVE

a) RED DE DEPENDENCIAS



b) PERFILES DE ATRIBUTO



c) CARACTERÍSTICAS DEL ATRIBUTO

Modelo de minería: Docente Visor: Visor Bayes naive de Microsoft

Red de dependencias Perfiles del atributo **Características del atributo** Distinción del atributo

Atributo: Hrs# _ Capacitación Valor: < 213,9063039488

Características para < 213,9063039488

Atributos	Valores	Probabilidad
Hrs# _ Tardanza	< 1,7685573676	
Experiencia	< 2	

mouse pointer inside or press Ctrl+G.

d) DISTINCIÓN DEL ATRIBUTO

Modelo de minería: Docente Visor: Visor Bayes naive de Microsoft

Red de dependencias Perfiles del atributo **Características del atributo** **Distinción del atributo**

Atributo: Hrs# _ Capacitación Valor 1: < 213,9063039488 Valor 2: Todos los otros esta

Puntuaciones de distinción para < 213,9063039488 y Todos los otros estados

Atributos	Valores	Favorece < 213,9063039...	Favorece Todos los otros ...
Hrs# _ Tardanza	< 1,7685573676		
Experiencia	< 2		
Hrs# _ Tardanza	1,7685573676 - 3,3333333332		

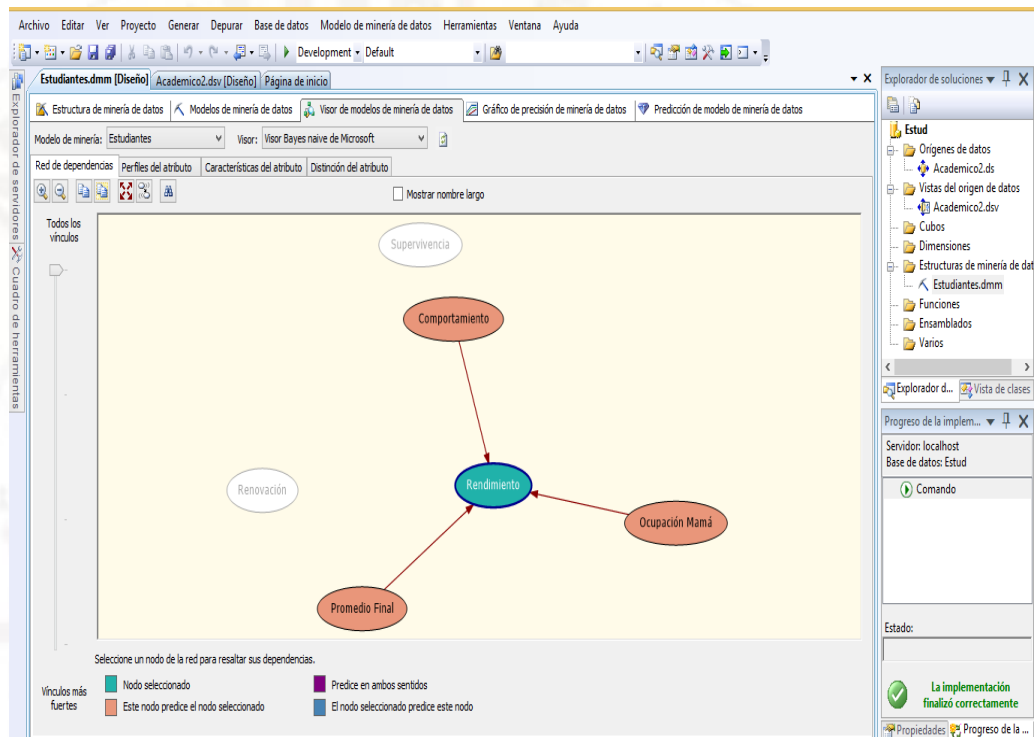
Leyenda de minería de datos

	< 213,9063039488	Todos los otros
Hrs# _ Tardanza = ...	2	0
Hrs# _ Tardanza !...	0	15

mouse pointer inside or press Ctrl+G.

4. MODELO DE MINERÍA DE DATOS DE ESTUDIANTES: VISOR BAYES NAIVES

a) RED DE DEPENDENCIAS

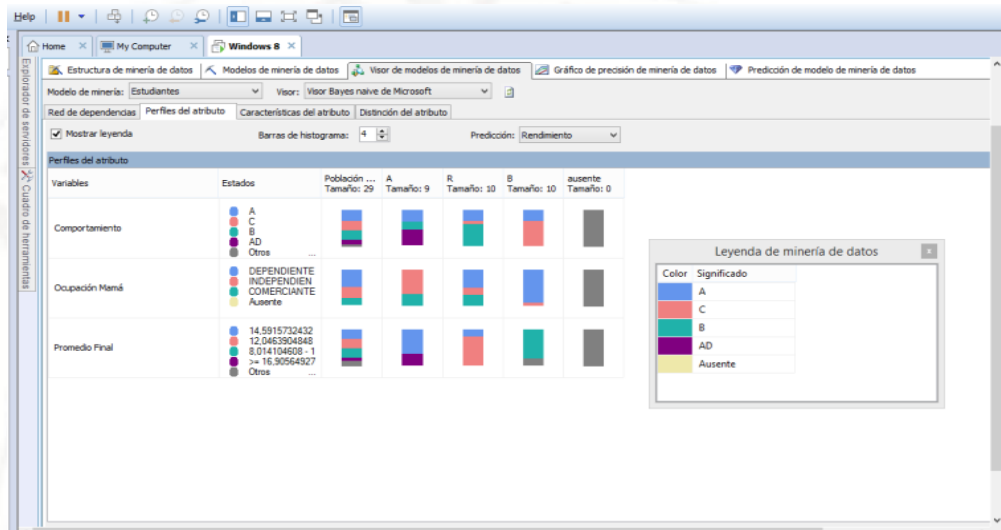


b) PERFILES DE ATRIBUTO

Atributo: Rendimiento Valor: A

Atributos	Valores	Probabilidad
Ocupación Mamá	INDEPENDIENTE	
Promedio Final	14,5915732452 - 16,9056492768	
Comportamiento	AD	
Ocupación Mamá	COMERCIANTE	
Promedio Final	>= 16,9056492768	
Comportamiento	A	
Comportamiento	B	

c) CARACTERISTICAS DEL ATRIBUTO



d) DISTINCIÓN DEL ATRIBUTO

