

Universidad Nacional
Federico Villarreal

Vicerrectorado de
INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POSGRADO

**"METODOLOGÍA DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE
TOMA DE DECISIONES DEL RENDIMIENTO ACADÉMICO DE LA
UNIVERSIDAD NACIONAL DE CAÑETE"**

**TESIS PARA OPTAR EL GRADO ACADEMICO DE
MAESTRO EN INGENIERIA DE SISTEMAS CON MENCIÓN EN INGENIERIA
DE SOFTWARE**

AUTOR:

RICARDO CARLOS INQUILLA QUISPE

ASESOR:

DR. JORGE VICTOR MAYHUASCA GUERRA

JURADO:

DR. SANTIAGO ESTEBAN CONTRERAS ARANDA

DR. PEDRO MARTIN LEZAMA GONZALES

M.SC. JESUS ELIAS CARRILLO BALCEDA

LIMA – PERU

2019

DEDICATORIA

El presente trabajo va dedicado a mis padres Miguel Inquilla Choqueña, Fermina Quispe Quispe y mi hermana Luz Inquilla Quispe, quienes me han siempre deseado mi bienestar en mi vida. Ellos representan al Dios Creador de todos los Sistemas en el Universo Expansivo y Pensante.

AGRADECIMIENTO

En Primer lugar, al Dios Todopoderoso creador de todo cuanto existe utilizando sus matemáticas espirituales me puso a mis padres para que siempre me den un aliento en esta vida. A mis amigos y compañeros de la Escuela Universitaria que me ha apoyado y dado aliento en momentos críticos.

INDICE

CARATULA.....	i
DEDICATORIAS.....	ii
AGRADECIMIENTOS.....	iii
INDICE.....	iv
ÍNDICE DE FIGURAS.....	vii
ÍNDICE DE TABLAS.....	viii
RESUMEN.....	ix
ABSTRACT.....	x
INTRODUCCIÓN.....	xi
1. CAPITULO I: PLANTEAMIENTO DEL PROBLEMA.....	15
2.1 DESCRIPCIÓN DEL PROBLEMA.....	15
1.2 FORMULACIÓN DEL PROBLEMA.....	22
1.2.1 PROBLEMA GENERAL.....	22
1.2.2 PROBLEMA ESPECIFICO.....	22
1.3 JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN.....	22
1.3.1 JUSTIFICACIÓN.....	22
1.3.2 AMBIENTAL.....	23
1.4 LIMITACIONES DE LA INVESTIGACIÓN.....	23
1.5 OBJETIVOS.....	23
1.5.1 OBJETIVO GENERAL.....	23
1.5.2 OBJETIVO ESPECIFICO.....	24
2. CAPITULO II: MARCO TEÓRICO.....	25
2.1 ANTECEDENTES.....	25
2.2.1 ANTECEDENTES INTERNACIONALES.....	25
2.2.2 ANTECEDENTES NACIONALES.....	30
2.2 MARCO CONCEPTUAL.....	33
2.3 TOMA DE DESICIONES.....	33

2.3.1 ETAPAS EN LA TOMA DE TOMA DESICIONES _____	33
2.3.2 TIPOS DE TOMA DE DESICIONES _____	34
2.4 INTELIGENCIA DE NEGOCIOS _____	36
2.5 SISTEMAS DE SOPORTE DE TOMA DE DESICIONES _____	37
2.5.1 SISTEMAS OLTP _____	38
2.5.2 SISTEMAS OLAP _____	38
2.5.3 DATAWAREHOUSE _____	39
2.5.4 DATAMART _____	40
2.5.5 METODOLOGIAS DE INTELIGENCIA DE NEGOCIOS _____	40
2.5.6 CUADRO COMPARATIVO DE METODOLOGIAS RELEVANTES _____	44
2.5.7 PENTAHO _____	44
2.5.8 MINERIA DE DATOS _____	47
2.5.9 RENDIMIENTO ACADEMICO _____	56
2.6 ASPECTOS DE RESPONSABILIDAD SOCIAL Y MEDIO AMBIENTE _____	56
3. CAPITULO III: METODO _____	57
3.1 TIPO Y NIVEL DE INVESTIGACIÓN _____	57
3.1.1 TIPO DE INVESTIGACIÓN _____	57
3.1.2 NIVEL DE INVESTGACIÓN _____	57
3.1.3 DISEÑO DE INVESTIGACIÓN _____	57
3.2 POBLACIÓN Y MUESTRA _____	58
3.2.1 POBLACIÓN _____	58
3.2.2 MUESTRA _____	58
3.3 HIPOTESIS _____	58
3.3.1 HIPOTESIS GENERAL: _____	58
3.3.2 HIPOTESIS ESPECÍFICAS: _____	58
3.4 OPERACIONALIZACIÓN DE VARIABLES _____	59
3.4.1 VARIABLE INDEPENDIENTE: _____	59
3.4.2 VARIABLE DEPENDIENTE: _____	60
3.5 INSTRUMENTOS _____	61
3.5.1 TECNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN DE CAMPO _____	61
3.5.2 TÉCNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN EXPERIMENTAL _____	61
3.5.3 TÉCNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN DOCUMENTAL _____	62
3.6 PROCEDIMIENTOS _____	62

3.6.1 ESTUDIO DE PLANIFICACIÓN	62
3.6.2 PLANIFICACIÓN DEL PROYECTO	63
3.6.3 DEFINICIÓN DE REQUERIMIENTOS DEL NEGOCIOS DEL BI_DM	67
3.6.4 DISEÑO DE LA ARQUITECTURA TECNICA DEL BI_DM	68
3.6.5 SELECCIÓN DE PRODUCTOS E IMPLEMENTACIÓN	69
3.6.6 MODELADO DIMENSIONAL	69
3.6.7 DISEÑO FISICO	73
3.6.8 DISEÑO E IMPLEMENTACIÓN DEL SUBSISTEMA DE ETL	78
3.6.9 ESPECIFICACIÓN DE APLICACIONES BI	80
3.6.10 DESARROLLO DE APLICACIONES DE BI	81
3.6.11 IMPLEMENTACIÓN	83
3.6.12 MINERIA DE DATOS	87
3.7 ANÁLISIS DE DATOS	93
3.7.1 RESULTADOS GENERICOS	93
3.7.2 RESULTADOS ESPECIFICOS	95
3.7.3 RESULTADOS NUMÉRICOS	96
4. CAPITULO IV: RESULTADOS	97
4.2 CÓNTRASTACIÓN DE HIPOTESIS	97
4.3 ANÁLISIS E INTERPRETACIÓN DE LA ESTADISTICA DESCRIPTIVA	101
5. DISCUSIÓN DE RESULTADOS	106
5.1 DISCUSION	106
5.2 CONCLUSIONES	107
5.3 RECOMENDACIONES	108
6. REFERENCIAS	109
7. ANEXOS	111

INDICE DE IMÁGENES

FIGURA 1. UBICACIÓN DE LA SEDE ADMINISTRATIVA DE LA UNIVERSIDAD NACIONAL DE CAÑETE	16
FIGURA 2. ORGANIGRAMA DE LA UNIVERSIDAD NACIONAL DE CAÑETE	18
FIGURA 3. PROCESO DE TOMA DE DECISIONES DEL RENDIMIENTO ACADÉMICO (AS-IS).....	20
FIGURA 4. PROCESO DE TOMA DE DECISIONES DEL RENDIMIENTO ACADÉMICO (TO-BE)	21
FIGURA 5: ETAPAS EN LA TOMA DE DESICIONES	34
FIGURA 6: TIPOS DE DECISIONES	35
FIGURA 7: LA INTELIGENCIA DE NEGOCIOS	36
FIGURA 8: ACTIVIDADES DE LA METODOLOGÍA EFESTO	41
FIGURA 9: DATAWARE HOUSE COORPORATIVO DE BILL INMON.....	42
FIGURA 10: METODOLOGÍA DE RALPH KIMBALL.....	43
FIGURA 11: LOGO DEL SOFTWARE PENTAHO	45
FIGURA 12: ARQUITECTURA DE PENTAHO	46
FIGURA 13: PROCESO DE CREACIÓN Y VALIDACIÓN DE UN MODELO BASADO EN APRENDIZAJE SUPERVISADO	50
FIGURA 14: PARTES DE UNA RED NEURONAL BIOLÓGICA.....	52
FIGURA 15: MODELO MATEMÁTICO DE UNA RED NEURONAL ARTIFICIAL SENCILLA.....	52
FIGURA 16: (A) RED PERCEPTRÓN CON 5 ENTRADAS Y 3 SALIDAS. (B) UN GRAFICO DE SALIDA DEL PERCEPTRÓN DE 2 UNIDADES. 53	
FIGURA 17: RED NEURONAL ARTIFICIAL CON UNA CAPA OCULTA Y 10 ENTRADAS.....	53
FIGURA 18: SUPERVIVENCIA DE LOS EMBARCADOS EN EL TITANIC	54
FIGURA 19: SERVIDOR PRO ELIANT DL 3000	66
FIGURA 20: DISEÑO DE LA ARQUITECTURA TÉCNICA.....	68
FIGURA 21: MODELO TRANSACCIONAL	73
FIGURA 22: MODELO ESTRELLA	74
FIGURA 23: MODELO DE CARGA DE DIMENSIONES Y TABLA DE HECHOS	79
FIGURA 24: MODELO DE DISEÑO DE CUBO	80
FIGURA 25: INTERFAZ DEL JPIVOT.....	81
FIGURA 26: PAGINA INICIAL DE PENTAHO	83
FIGURA 27: ASIGNATURAS MATRICULADAS	84
FIGURA 28: ASIGNATURAS MATRICULADAS EN BARRA	84
FIGURA 29: ASIGNATURAS APROBADAS EN BARRAS MÚLTIPLES	85
FIGURA 30: ASIGNATURAS DESAPROBADAS POR SEMESTRE ACADÉMICO.....	85
FIGURA 31: CRÉDITOS MATRICULADOS POR SEMESTRE	86
FIGURA 32: CRÉDITOS APROBADOS POR SEMESTRE ACADÉMICO	86
FIGURA 33: CRÉDITOS DESAPROBADOS POR SEMESTRE	87
FIGURA 34: DESCRIPCIÓN EXPLORATIVA DE LOS ATRIBUTOS	88
FIGURA 35: RESULTADOS CON WEKA.....	89

<i>FIGURA 36: MATRIZ DE CONFUSIÓN</i>	89
<i>FIGURA 37: VISUALIZACIÓN DEL ÁRBOL EN MODO TEXTUAL</i>	90
<i>FIGURA 38: VISUALIZACIÓN DEL ÁRBOL</i>	90
<i>FIGURA 39: MATRIZ DE CONFUSIÓN</i>	91
<i>FIGURA 40: MUESTRA DE PREDICCIONES DEL MODELO</i>	92
<i>FIGURA 41: CRITERIO DE DECISIÓN DE KPI₁</i>	98
<i>FIGURA 42: DISTRIBUCIÓN DE PROBABILIDAD DEL KPI₂</i>	100
<i>FIGURA 43: RESUMEN DE POST-PRUEBA O₂ PARA KPI 1.</i>	101
<i>FIGURA 44: RESUMEN PARA KPI 2 DE POST-PRUEBA O₂</i>	102
<i>FIGURA 45: RESUMEN DE PRE-PRUEBA O₁ PARA KPI 3.</i>	104
<i>FIGURA 46: RESUMEN DE POST-PRUEBA O₂ PARA KPI 3.</i>	105

INDICE DE TABLAS

TABLA 1: DATOS ACTUALES DE LOS INDICADORES	19
TABLA 2: TABLA COMPARATIVA MBALL VS INMON	44
TABLA 3: TIPOLOGÍA DE LOS ALGORITMOS	50
TABLA 4: OPERACIONALIZACIÓN DE VARIABLE INDEPENDIENTE	59
TABLA 5: OPERACIONALIZACIÓN DE VARIABLE DEPENDIENTE	60
TABLA 6: TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN DE CAMPO.	61
TABLA 7: TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN EXPERIMENTAL	61
TABLA 8: TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN DOCUMENTAL.	62
TABLA 9: TÉCNICAS E INSTRUMENTOS DE LA INVESTIGACIÓN EXPERIMENTAL	65
TABLA 10: ESPECIFICACIONES TÉCNICAS DEL SERVIDOR PRO ELIANT DL 3000	66
TABLA 11: ESPECIFICACIONES TÉCNICAS DE PCS DE USUARIOS	66
TABLA 12: LISTA DE SOFTWARE.....	67
TABLA 13: LISTA DE REQUERIMIENTOS	67
TABLA 14: ROL DE LOS PRODUCTOS DE SOFTWARE.....	69
TABLA 15: MEDIDAS Y FORMULAS	69
TABLA 16: DIMENSIONES Y MEDIDAS.....	70
TABLA 17: DIMENSIÓN ALUMNO.....	70
TABLA 18: DIMENSIÓN ASIGNATURA.....	71
TABLA 19: DIMENSIÓN CARRERA.....	71
TABLA 20: DIMENSIÓN CICLO	71
TABLA 21: DIMENSIÓN PERIODO	72
TABLA 22: TABLA DE HECHOS.....	72
TABLA 23: TABLA DE MEDIDAS	72
TABLA 24: DICCIONARIO DE DATOS DE LA DIMENSIÓN ALUMNO	75
TABLA 25: DICCIONARIO DE DATOS DE LA DIMENSIÓN ASIGNATURA.....	76
TABLA 26: DICCIONARIO DE DATOS DE LA DIMENSIÓN CARRERA	76
TABLA 27: DICCIONARIO DE DATOS DE LA DIMENSIÓN CICLO	77
TABLA 28: DICCIONARIO DE DATOS DE LA DIMENSIÓN PERIODO	77
TABLA 29: CARGA DE DIMENSIONE	78
TABLA 30: CARGA DE TABLA DE HECHOS.....	79
TABLA 31: DICCIONARIO DE DATOS DE RECOPIACIÓN DE LA INFORMACIÓN	88
TABLA N°32: REGISTRO DE DATOS DE PRE-PRUEBA O ₁ Y POST-PRUEBA O ₂ PARA LOS KPI ₁ , KPI ₂ , KPI ₃	95
TABLA N°33: PROMEDIO DE LOS INDICADORES O ₂ Y O ₁	96
TABLA N°34: RESUMEN DE PRUEBA T STUDENT DE KPI ₁	98
TABLA N°35: RESUMEN DE PRUEBA T STUDENT DE KPI ₂	100

TABLA N°38: VALORES DE PRE-PRUEBA O ₁ PARA KPI 3.....	103
TABLA N°39: VALORES DE POST-PRUEBA O ₂ PARA KPI 3.....	104
TABLA N°40: MATRIZ DE CONSISTENCIA.....	111
TABLA N°41: BIENES UTILIZADOS EN EL DESARROLLO DE LA TESIS.....	112
TABLA N°42: BIENES UTILIZADOS EN EL DESARROLLO DE LA TESIS.....	112
TABLA N°43: SERVICIOS UTILIZADOS EN EL DESARROLLO DE LA TESIS.....	113

RESUMEN

METODOLOGÍA DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE TOMA DE DECISIONES DEL RENDIMIENTO ACADÉMICA DE LA UNIVERSIDAD NACIONAL DE CAÑETE, 2018

Inquilla, Ricardo

rinquilla@undc.edu.pe

El proceso de toma de decisiones es una tarea estratégica por parte de los responsables de los procesos, en el ámbito académico supone la aplicación de actividades y estrategias que permitan optimizar el rendimiento académico en las instituciones educativas utilizando herramientas tecnológicas que les permitan obtener el conocimiento en un tiempo óptimo.

La presente tesis establecer una propuesta de metodología de inteligencia de negocios y minería de datos denominada BI_DM que permita dar soporte en el proceso de toma de decisiones, en la presente tesis se desarrolla del área académica de la Universidad Nacional de Cañete. La satisfacción de los usuarios debe ser el principal objetivo de las instituciones públicas, por ello los responsables de la administración pública deben contar con herramientas eficientes que doten de respuestas rápidas a problemáticas académicas que apoyen en políticas de bienestar a los alumnos. En este mundo tan vertiginoso que desea respuestas rápidas de información la Metodología de Inteligencia de Negocios es un aliado a esta creciente demanda de servicios en tiempo real y que respondan necesidades a problemáticas académicas.

El capítulo uno, describiremos el planteamiento del problema, delimitaciones del problema, objetivos generales y específicos y una breve justificación e importancia de la investigación.

Capítulo dos, se detallan bases teóricas de inteligencia de negocios y minería de datos para el desarrollo de la solución

Capítulo 3, se describe el uso de la metodología de inteligencia de negocios.

Capítulo 4, se presenta los resultados de la solución planteada y por último se detallan las conclusiones y recomendaciones a las que se ha llegado al realizar la investigación del proyecto.

Palabras Clave: Inteligencia de Negocios, Business Intelligence, Bigdata, Datamart, Pentaho, Minería de datos

ABSTRACT**METHODOLOGY OF BUSINESS INTELLIGENCE IN THE DECISION-
MAKING PROCESS OF THE ACADEMIC AREA OF THE NATIONAL
UNIVERSITY OF CAÑETE, 2018****Inquilla, Ricardo**

rinquilla@undc.edu.pe

The process of decision making is a strategic task on the part of those responsible for the processes, in the academic field it involves the application of activities and strategies that allow optimizing academic performance in educational institutions using tools that allow them to obtain knowledge in an optimal time.

The present thesis establish a proposal of business intelligence methodology and data mining called BI_DM that allows to support the decision making process, in this thesis is developed from the academic area of the National University of Cañete. The satisfaction of users should be the main objective of public institutions, therefore, those responsible for public administration should have efficient tools that provide rapid responses to academic problems that support welfare policies for students. In this fast-paced world that desires rapid information responses, the Business Intelligence Methodology is an ally to this growing demand for services in real time and that responds to the needs of academic problems.

Chapter one, we will describe the approach of the problem, delimitations of the problem, general and specific objectives and a brief justification and importance of the investigation.

Chapter two, theoretical bases of business intelligence and data mining for the development of the solution are detailed

Chapter 3, describes the use of business intelligence methodology.

Chapter 4, the results of the proposed solution are presented and finally the conclusions and recommendations that have been reached when carrying out the project research are detailed.

Keywords: Business Intelligence, Bigdata, Datamart Datawarehouse, Data Mining.

INTRODUCCIÓN

Los avances en la tecnología han llevado a generar nuevas soluciones de inteligencia de negocios que han pasado a formar parte de las herramientas para el proceso de toma de decisiones en las empresas e instituciones educativas que se necesitan con suma urgencia tomar decisiones teniendo información en los tiempos adecuados.

Con base en lo anterior, se necesita el desarrollo de nuevas soluciones de inteligencia de negocios pueda permitir mejorar el proceso de toma de decisiones teniendo en cuenta factores como el tiempo, la disminución de errores y aumentar la satisfacción de los tomadores de decisiones.

La dificultad de mejorar el proceso de toma de decisiones conlleva a proponer una serie de estrategias para solucionar el problema. La actual base de datos no nos permite obtener la información en los tiempos adecuados y de una forma amigable que aumente la experiencia del usuario en el uso de herramientas para mejorar el proceso de toma de decisiones.

La presente tesis propone la sistematización de la implementación de la metodología de inteligencia de negocios denominado BI_DM que permitan mejorar el proceso de toma de decisiones en el área académica de la Universidad Nacional de Cañete.

BI_DM es una metodología de desarrollo de Sistemas de Inteligencia de Negocios para el ámbito académico con el objetivo de dar soporte el proceso de Toma de Decisiones.

1. CAPITULO I: PLANTEAMIENTO DEL PROBLEMA

2.1 DESCRIPCIÓN DEL PROBLEMA

MUNDIAL

En las grandes Organizaciones modernas se administra un gran flujo de información y sobre todo al ritmo que vamos tenemos diferentes requerimientos de información para tomar decisiones, el reto está en tener herramientas que nos permitan tomar decisiones en los negocios. Gracias a los adelantos en las Tecnologías de la Información tenemos la posibilidad de recolectar datos de manera impresionante. Basta con observar la información que recolecta un centro comercial por medio de las tarjetas, por una cámara de seguridad en las calles del mundo. Lamentablemente para muchas empresas, el tener enormes cantidades de datos se ha convertido en un problema al no tener herramientas que nos permitan obtener información relevante. Es en este punto que las herramientas de Inteligencia de Negocios nos permiten nos permite obtener conocimiento de los datos, con la intención de poder utilizarlos de una forma sencilla y entender el porqué de nuestro desempeño o, mejor aún, plantear escenarios a futuro, lo cual nos apoyara a seleccionar mejores decisiones. La tecnología de Inteligencia de Negocios resulta entonces sumamente importante. Podemos encontrar soluciones de BI con tecnología abierta y enfocadas al usuario final, no necesitamos considerar un gasto excesivo en aumentar nuestra área de experiencia. Estas soluciones están desarrolladas para que los tomadores de decisiones puedan realizarlo sin la intervención del área de TI, lo cual agiliza los procesos y permite una reacción inmediata a cualquier movimiento dentro de la industria o del mercado. La experiencia es la que irá conduciendo a la mayoría de los usuarios comunes hacia cada vez mejores resultados.

PERÚ

Con la recién aprobada Ley Universitaria 30220 y las Condiciones Básicas de Calidad de SUNEDU, supone que las Universidad deben manejar información que les permita obtener información para una toma de decisiones correcta para la implementación de mejores políticas

orientadas a la buena atención del servicio Educativo, por ello la gran cantidad de datos dispersos en las diferentes oficinas, se pueden centralizar en una herramienta de Inteligencia de Negocios, obteniendo buenos reportes e indicadores e indicadores que permitan un manejo y toma de decisiones que apoyen en los procesos académicos donde se pretende optimizar tiempos y aumentar la satisfacción de los estudiantes.

UNIVERSIDAD NACIONAL DE CAÑETE

La Universidad Nacional de Cañete, fue creada mediante Ley N° 29488, el veintidós de diciembre del 2009, con domicilio en el distrito de San Vicente, provincia de Cañete, departamento de Lima, sobre la base de las actuales sedes en dicha provincia de las Universidad Nacionales del Callao y José Faustino Sánchez Carrión; rige su actividades y funcionamiento por las disposiciones legales de la Constitución Política del Estado, la Ley Universitaria, su Estatuto, su Reglamento General y los demás Reglamentos específicos y disposiciones legales aplicables a las personas jurídicas de derecho público interno.



Figura 1. Ubicación de la Sede Administrativa de la Universidad Nacional de Cañete

El proceso académico es un proceso complejo con muchas aristas que debemos analizar para dar un buen servicio educativo, en este contexto la Universidad Nacional de Cañete ha venido ejecutando el proceso académico sin tener información de indicadores académicos relevantes para la toma de decisiones a nivel gerencial, actualmente la información se solicita en formato Excel o Word, el cual toma un tiempo excesivo la generación de reportes, además de ello no permite mostrar la información de manera amigable.

La situación actual no facilita a los distintos actores que participan en el proceso académico una toma de decisiones oportunas por lo que no se tiene información confiable y actualizada y los reportes académicos actuales no permiten filtros ni selección de opciones del semestre anterior para su análisis de su evolución en el tiempo.

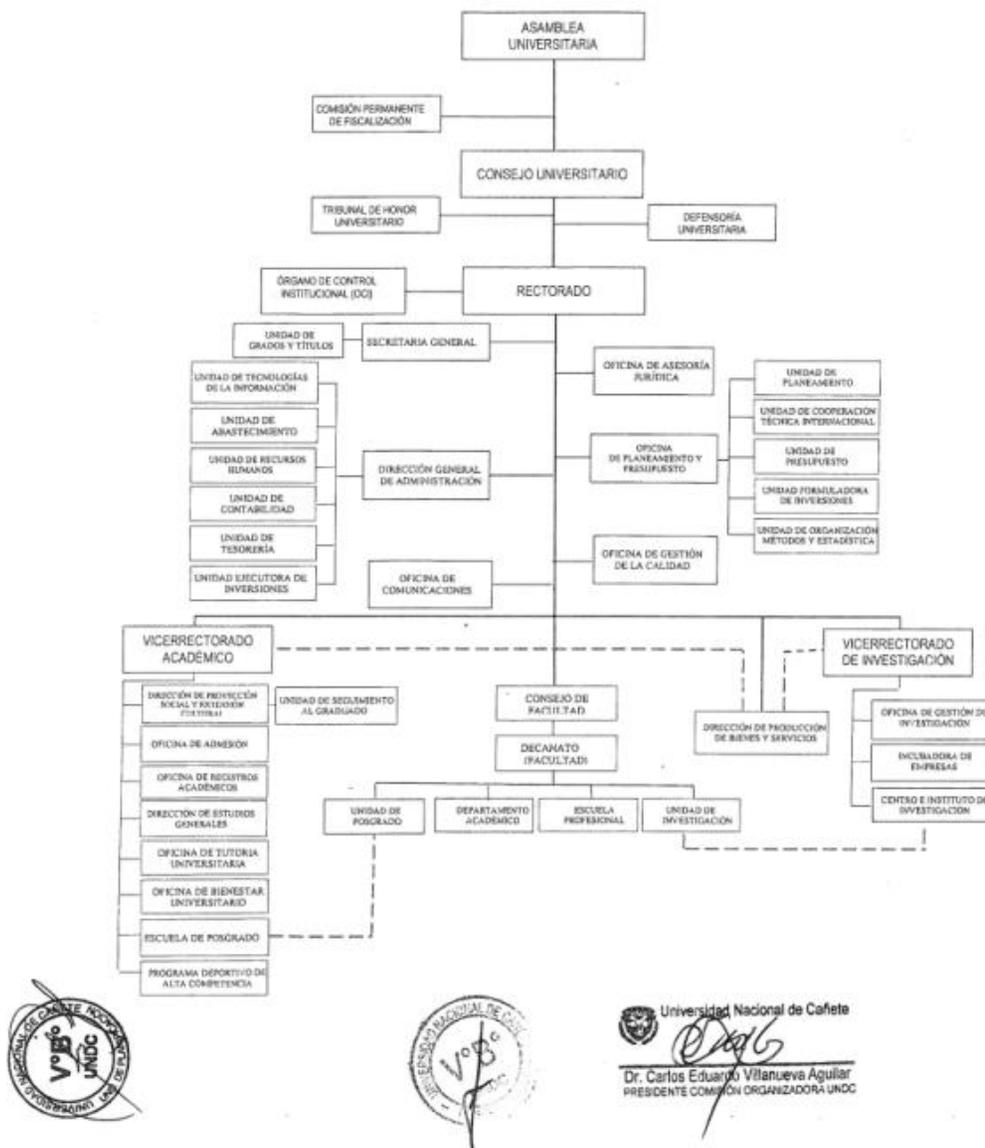


Figura 2. Organigrama de la Universidad Nacional de Cañete

Además de ello no existe un conjunto de cuadro de mando de indicadores del proceso académico que permite tomar decisiones con el fin de mejorar el proceso académico.

El proceso muestra los siguientes problemas en

- Tiempo excesivo en la generación reportes
- Tiempo excesivo en el análisis de la información
- Bajo nivel de satisfacción de los usuarios

TABLA 1:
Datos actuales de los indicadores

Indicadores (KPI)	Valores
Tiempo excesivo en la generación reportes	84 segundos
Tiempo excesivo en el análisis de la información	90 minutos
Bajo nivel de satisfacción de los usuarios	Malo

Fuente: Elaboración Propia

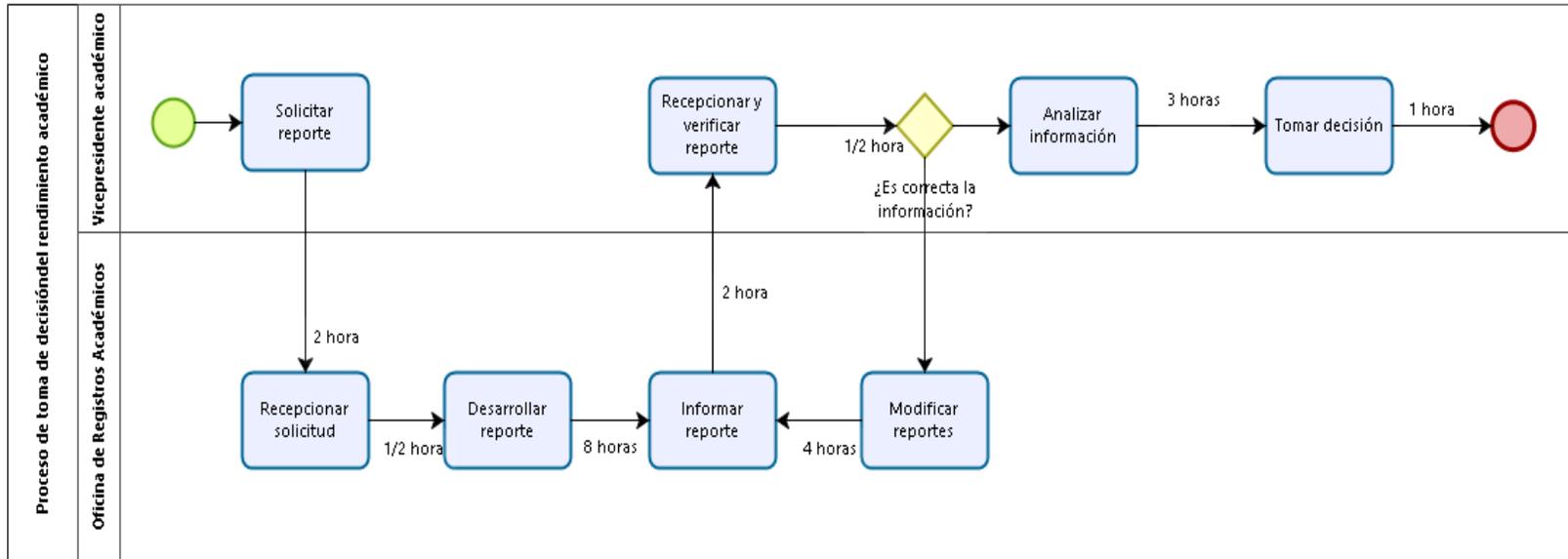


Figura 3. Proceso de toma de decisiones del rendimiento académico (AS-IS)

De acuerdo a la Figura 3, se propone realizar una mejora al Proceso de toma de decisiones quedando de la siguiente

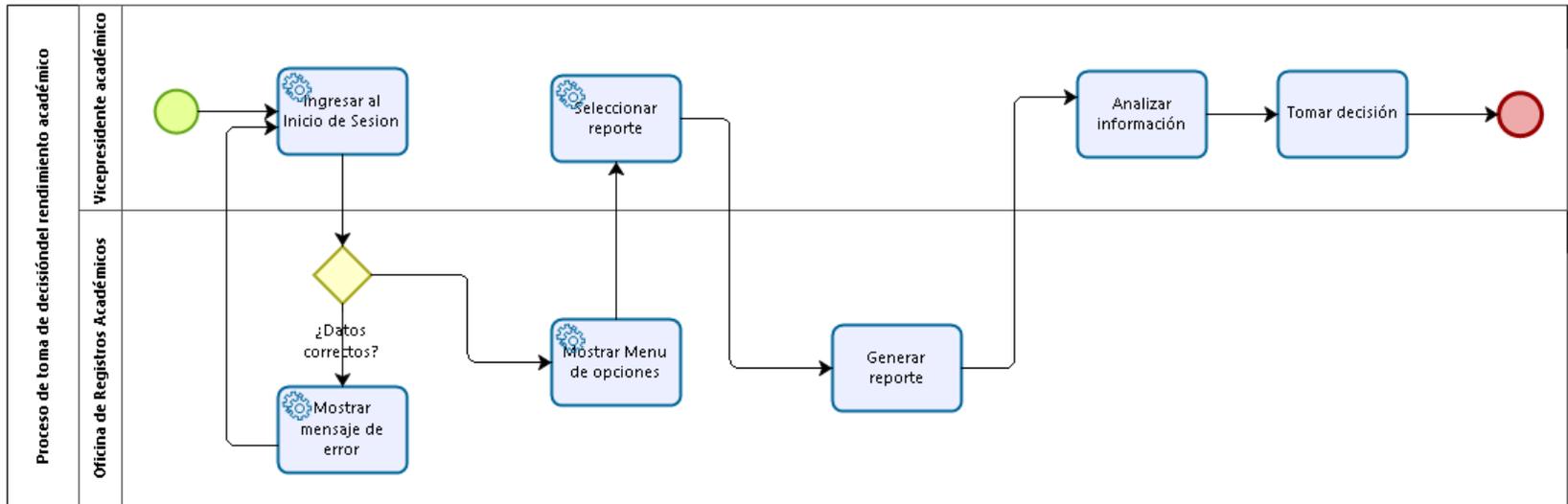


Figura 4. Proceso de toma de decisiones del rendimiento académico (TO-BE)

1.2 FORMULACIÓN DEL PROBLEMA

1.2.1 PROBLEMA GENERAL

¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete?

1.2.2 PROBLEMA ESPECIFICO

- A. ¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el tiempo de generación de reportes del proceso de toma de decisiones del rendimiento académico?
- B. ¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el tiempo de análisis de la información del proceso de toma de decisiones del rendimiento académico?
- C. ¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en la satisfacción del usuario del proceso de toma de decisiones del rendimiento académico?

1.3 JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN

1.3.1 JUSTIFICACIÓN

1.3.1.1 RELEVANCIA SOCIAL

El presente proyecto permite proponer de una Metodología de Inteligencia de Negocios, denominado “BI_MD” que permite optimizar el Proceso de toma de decisiones acerca del rendimiento lo cual beneficiará en mejores decisiones para el servicio educativo de los estudiantes de la Universidad Nacional de Cañete

1.3.1.2 TEÓRICA

La presente investigación permite incrementar los conocimientos del proceso de toma de decisiones, lo cual permitirá servir de base para futuras investigaciones.

1.3.1.3 PRÁCTICA

La implementación de la metodología de inteligencia de negocios denominada BI_DM permitirá utilizar eficaz y eficientemente la infraestructura tecnológica de la Intranet de la Universidad Nacional de Cañete. Además de ello nos permite cumplir con la ley Universitaria 30220 (Art. 11) acerca de la transparencia de las universidad así como el cumplir la CONDICION VIII : indicador 55 del proceso de Licenciamiento de las Universidad de acuerdo a la Resolución 54 -2017-SUNEDU

1.3.2 AMBIENTAL

El impacto ambiental del presente proyecto se justifica en el mejor manejo de recursos tanto en el aspecto de ahorro de energía, consumo de papel, impresiones y demás materiales consumidos en la generación de informes pertinentes para la toma de decisiones.

1.3.2.1 IMPORTANCIA

La decisión de implementar metodología de inteligencia de negocios denominada “BI_DM” es importante porque lo que se pretende es optimizar el proceso de Toma de decisiones académicas, permitiendo a la Universidad Nacional de Cañete una eficaz y eficiente gestión académica, y logrando elevar la satisfacción de los alumnos, docentes y autoridades.

1.4 LIMITACIONES DE LA INVESTIGACIÓN

No se ha encontrado limitaciones sustanciales que no permitan desarrollar la presente investigación debido a que los requerimientos de información, eran urgente para cumplir con las leyes vigentes de la nueva Ley Universitaria 30220 y la Resolución 054-SUNEDU.

1.5 OBJETIVOS

1.5.1 OBJETIVO GENERAL

Determinar el grado de influencia de la metodología de inteligencia de negocios en el proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete.

1.5.2 OBJETIVO ESPECIFICO

- A. Determinar el grado de influencia de la de metodología de inteligencia de negocios en el tiempo de generación de reporte del proceso de toma de decisiones del rendimiento académico.

- B. Determinar el grado de influencia de la metodología de metodología de inteligencia de negocios en el tiempo de análisis de la información del proceso de toma de decisiones del rendimiento académico.

- C. Determinar el grado de influencia de la metodología de inteligencia de negocios en el nivel de satisfacción del usuario del proceso de toma de decisiones del rendimiento académico.

2. CAPITULO II: MARCO TEÓRICO

2.1 ANTECEDENTES

2.2.1 ANTECEDENTES INTERNACIONALES

Bakar, M., Ta'a, A & Soid, M.(2018). En su artículo “Data warehouse system for blended learning in institutions of higher education. e-Academia Journal”, se tuvo como objetivo desarrollar un modelo DW para monitorear el uso de aprendizaje combinando el e-learning (encuentros asincrónicos) con encuentros presenciales (sincrónicos) posteriormente, por lo que esta investigación desarrolló un modelo de análisis de datos para monitorear el uso del e-learning combinado, especialmente el uso de la tecnología de aprendizaje combinado por parte de los docentes. La metodología consistió en el proceso de limpiar y transformar datos hacia los esquemas del Data Warehouse particular. La solución permitió a la gerencia y a los docentes consultar el estado del aprendizaje.

DeSá Mussa, M., de Souza, S. C. & da Silva Freire, E. F (2018), en su artículo científico denominado “BUSINESS INTELLIGENCE IN EDUCATION: AN APPLICATION OF PENTAHO SOFTWARE”, afirma que “los sistemas de Business Intelligence (BI) se pueden definir como para asistir y extender los procesos de toma de decisiones y hacerlos más precisos y confiables, basados en la conocimiento generado por los datos de la empresa”, por ello implementan la solución con el Software Pentaho en el contexto de una institución educativa, a través de la creación de un modelo multidimensional y la inserción de datos académicos de un centro educativo, lograron tener con mayor rapidez los reportes de manera rápida que podría ser utilizada por los gerentes y la toma de desiciones relvantes del proceso académico.

Acosta, La Red Martínez & Primorac (2018), en su artículo denominado: “Determinación de perfiles de rendimiento académico en la UNNE con Minería de Datos Educativa” afirma

“Evaluar el rendimiento de los estudiantes mediante técnicas de Minería de Datos. La propuesta no se enfoca en analizar el perfil del estudiante solo a través de sus calificaciones, sino también, estudiar el desempeño académico en base a otras variables. Para definir los perfiles de los estudiantes y determinar patrones que conduzcan al éxito o fracaso académico, implementaremos un modelo que relaciona las calificaciones de los estudiantes con otras variables, tales como factores socioeconómicos, demográficos, actitudinales, entre otros; en base a lo cual clasificaremos los diferentes perfiles de alumnos”.

Zambrano (2017), en su TR5QAtesis: “Implementación de una solución de inteligencia de negocios acerca de la información de los docentes, estudiantes y personal administrativo de la Universidad Técnica del Norte para el Instituto de Altos Estudios” plantea:

“Una solución de Inteligencia de Negocios para el análisis de la información tanto de los docentes como de los estudiantes de la Universidad Técnica del Norte, con el propósito de servir como un indicador para la mejor toma de decisiones por parte del Instituto de Altos Estudios concernientes a sus respectivas investigaciones”.

Arenas, Maria y Gomez, Ana (2017), en su tesis: “Inteligencia de negocios aplicada a los procesos de autoevaluación de la Universidad de Manizales”, plantea:

“Una metodología de inteligencia de negocios a los procesos de autoevaluación con el de identificar la necesidad de que estos sean analizados de forma adecuada para extraer información y conseguir que sean visualizados por medio de reportes e indicadores de forma gráfica para tomar decisiones o simplemente identificar factores que deban ser mejorados a partir de las evaluaciones registradas”.

Arenas y Gómez (2017), en su artículo denominado “Proceso de Explotación de Información para Analítica Académica en FaCENA-UNNE”, dice:

“La explotación de la información académica es cada vez más necesaria. La analítica académica combina los datos institucionales, el análisis estadístico y los modelos predictivos permitiendo la exploración de datos para identificar informaciones nuevas y útiles para atender

las expectativas y necesidades estratégicas de las organizaciones de educación superior. Academic analytics o Analítica académica es un nuevo campo surgido en la educación superior como consecuencia de las prácticas de minería de datos y la utilización de herramientas de inteligencia de negocios. Puede referirse ampliamente a las prácticas de toma de decisiones basadas en datos para fines operativos a nivel de universidad, pero también puede ser aplicado a las dificultades del proceso de enseñanza y aprendizaje de los estudiantes. Por ejemplo, predecir la probabilidad de abandono de los estudiantes o el tiempo de finalización de los estudios, aunque en la actualidad, el énfasis está puesto en inteligencia procesable, información que puede ser entregada con tiempo suficiente para hacer una diferencia en el rendimiento académico”.

Holguín y Monserrate (2015), en su tesis: “Análisis, Diseño y Desarrollo del DataMart actividades de docentes y estudiantes para el prototipo de Sistema académico integrado en las carreras de Ingeniería en Sistemas Computacionales e Ingeniería en Networking y Telecomunicaciones de la Facultad de Ciencias Matemáticas y Física de la Universidad de Guayaquil”, afirma:

“La generación de reportes de indicadores con la finalidad de autoevaluar las carreras antes de su proceso de acreditación. Indica que para la elaboración de los indicadores se realiza de forma manual, lo cual demanda una gran cantidad de tiempo para su obtención. Esta tesis propone el Desarrollo del Datamart Actividades Docentes y Estudiantes para generar reportería de forma automática de los indicadores y control interno de las carreras. La inteligencia de negocios aplicada a las IES como una alternativa tecnológica, ayudará a manejar grandes volúmenes de datos históricos y convertirlos en conocimientos para los directivos, sirviendo de apoyo”.

Giovaninni, Karanik, Pinto y La Red Martínez (2015), en su artículo titulado: “Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos”, afirma:

“El rendimiento académico como un factor crítico teniendo en cuenta que, frecuentemente, el bajo rendimiento académico está asociado a una alta tasa de deserción. Esto se ha observado

en asignaturas del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la Universidad Tecnológica nacional facultad Regional Resistencia (UTn-fRRe), situada en la ciudad de Resistencia, provincia del Chaco, Argentina, entre ellas Algoritmos y Estructura de datos, donde el bajo rendimiento académico se observa en proporciones muy altas (entre el 60% y el 80% aproximadamente en los últimos años). En este trabajo se propone la utilización de técnicas de minería de datos sobre información del desempeño de los alumnos de la asignatura mencionada con el propósito de caracterizar los perfiles de alumnos exitosos (buen rendimiento académico) y de aquellos que no lo son (bajo rendimiento académico). La determinación de estos perfiles permitiría a futuro definir acciones específicas tendientes a revertir el bajo rendimiento académico, una vez detectadas las variables asociadas al mismo. En este artículo se describen los modelos de datos y de minería de datos utilizados y se comentan los principales resultados obtenidos”.

Vázquez, Frías, González y Farías (2018), en su artículo: “Análisis de información de aspirantes a ingresar a instituciones de educación superior utilizando Pentaho” realiza:

“Un estudio sobre la información recabada en base de datos mediante el registro y validación de aspirantes a la Universidad, debido a la naturaleza del proceso, la recopilación y manejo de datos, dicha base de datos es consistente en datos de rendimiento académico, entre las que se mencionan: resultado de evaluaciones, becas, entre otros, sin perder de vista y tomando en consideración las políticas institucionales de manejo de información y datos personales, se identifica el procesamiento como un área de oportunidad para las entidades de desarrollo académico y seguimiento curricular de dichas instituciones. Las áreas de impacto que se identifican dentro de este proceso es la toma de decisiones en base a datos estadísticos sobre deserción escolar, análisis de escuelas de procedencia, fortalezas y debilidades en áreas de conocimiento, programas de tutoría y seguimiento académico. El presente trabajo considera una propuesta de análisis utilizando las tecnologías de información y comunicación aplicada en la educación, en el ámbito de inteligencia de negocios utilizando Pentaho como herramienta de procesamiento y obtención de datos estadísticos, basados en procesos y actividades concretos, de donde proviene la propuesta de implementar una alternativa de apoyo a la toma de decisiones enfocada a los departamentos o áreas de las IES”.

Pascal, Servetto, Unelén y Yésica (2017), en su artículo “Aplicación de Business Intelligence para la toma de decisiones en Instituciones Universitarias. Implementación de Boletines Estadísticos en la Universidad Nacional de Lomas de Zamora (UNLZ)”, afirma:

“Los resultados de la implementación de un Data Warehouse, una de las tecnologías que forman parte del Business Intelligence (BI), para la toma de decisiones estratégicas en organizaciones universitarias. El BI, consiste en el proceso de recopilación del capital intangible de una organización independientemente de la plataforma originada y la extracción de conocimiento según los requerimientos del usuario. Los avances corresponden a la implementación de boletines estadísticos diseñados para tres tipos de destinatarios, con una frecuencia preestablecida, que reflejan la situación actual de la institución”.

Santoyo (2018), en su tesis titulada: “Modelo de datamart para análisis de indicadores de desempeño estudiantil en la SEPI UPIICSA” escribe: “En el panorama educativo actual, determinadas carencias en materia de información y, sobre todo, las ausencias de indicadores eficientes que permitan evaluar a los egresados de las diferentes carreras que se imparten en Instituciones de Educación Superior, pueden repercutir en la retroalimentación para la mejora los procesos enseñanza-aprendizaje y mantener el vínculo con los mismos. En los últimos años las herramientas de Inteligencia de Negocio permiten analizar y explorar la información más relevante en una organización. El enfoque planteado en este trabajo se centra en el desarrollo de un Sistema de Información Web para Egresados, para la Sede Regional de Chiriquí de la Universidad Tecnológica de Panamá. Este sistema permite generar Dashboard donde se puede visualizar información útil para la toma de decisiones”.

Ayala y Omar (2018); en su tesis titulada: “Construcción de un Data Mart para la Gestión Académica de la Universidad Técnica del Norte(Bachelor's thesis).” Propone: “Apoyar la toma de decisiones que se realizan en el vicerrectorado académico de la Universidad Técnica del Norte, utilizando herramientas de inteligencia de negocios para el desarrollo del software necesario. Se ha optado por la mitología de Ralph Kimball para el diseño de la base de datos

multidimensional. El capítulo I presenta la idea que motivó el emprender este trabajo, en el capítulo II nos introduce en los conceptos de Inteligencia de Negocios. Las metodologías más populares para el desarrollo de un proyecto de data warehouse se detallan en el capítulo III, la implementación de los data mart se describe en el capítulo IV y finalmente en el capítulo V se redactan las conclusiones y recomendaciones”

2.2.2 ANTECEDENTES NACIONALES

Yamao (2018) en su tesis: “Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas, Universidad San Martín de Porres, Lima - Perú”, donde :

“Estudia la predicción del rendimiento académico de los alumnos que ingresaron a la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres en el primer ciclo utilizando minería de datos. Teniendo los siguientes objetivos:

Estimar indicadores sociales, económicos y académicos para el rendimiento académico en estudiantes universitarios de primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres.

Determinar mediante análisis de componentes principales la significación de los indicadores sociales, económicos y académicos en estudiantes universitarios de primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres.

Aplicar la minería de datos educacional para indicadores sociales, económicos y académicos en estudiantes universitarios de primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres”.

Rafael y Elizabeth (2017) en su tesis: “Implementación de un datamart para el seguimiento académico de los estudiantes en la Escuela académico profesional de Ingeniería de Sistemas de la Universidad Nacional de Cajamarca” realiza:

“Un estudio con el objetivo de realizar el seguimiento académico de los alumnos de la Escuela Académico Profesional de Ingeniería de Sistemas de la Universidad Nacional de Cajamarca y de esta manera poder apoyar en la toma de decisiones en cuanto a temas académicos con la implementación de un data mart bajo la metodología de Ralph Kimball. Se analizó el estado actual de la escuela, determinándose que, si bien se cuenta con un sistema proporcionado por la Unidad Técnica de Sistemas Informáticos de la Universidad Nacional de Cajamarca, este sistema presenta algunas limitaciones en cuanto a la información disponible, de la misma manera con el acceso a esta, aspecto que le demandan exceso de tiempo al momento de realizar el análisis de la información académica y por ende la toma de decisiones”.

Peñaloza y Diego (2015), en sus tesis: “Implementación de un Sistema de Soporte de Decisiones para la Gestión Académica de la Universidad Nacional José María Arguedas” presenta:

“Un trabajo donde brindó una solución de inteligencia de negocios que permitirá mejorar la gestión académica de la universidad. Este hecho se logró con la aplicación de la tecnología Datawarehouse como parte del sistema de información analítico para la gestión académica, que permitió obtener respuestas a las consultas requeridas de manera rápida y haciendo uso óptimo de los recursos. Para este fin se utilizó la metodología de Ralph Kimball que se ajusta más a lo que se quiere desarrollar al permitir la creación del Datawarehouse partiendo de los Datamart, al estar involucradas solamente las áreas académicas”.

Duran (2017), en su tesis titulada “DataMart como parte de una solución de inteligencia de negocios, para el soporte de la toma de decisiones de la gestión académica de la Facultad de Ciencias de la USAM”, realizó:

“Un estudio utilizando una metodología experimental, donde la cual la muestra de estudio estuvo constituida por los tomadores de decisiones de la Facultad de Ciencias, siendo el grupo experimental la Escuela de Ingeniería de Sistemas e Informática y el grupo control conformada por la Escuela de Estadística e Informática y la escuela de Matemática. Para la contratación de la hipótesis se utilizó la prueba de hipótesis de t-student donde cuya conclusión arribada fue, la implementación de un Data Mart, como parte de una solución de inteligencia de negocio, se

obtuvo una mejora significativamente el proceso de Toma de Decisiones en la Gestión Académica de Facultad la Ciencias de la UNASAM”.

Sarmiento (2018), en su tesis titulada “Inteligencia de negocios usando Pentaho para la gestión académica en la UNAMBA-2016”, realiza:

“Una investigación cuyo objetivo principal es es el implementar una propuesta de solución de inteligencia de negocios usando Pentaho, para contribuir a mejorar el análisis de la información en la Gestión Académica de la Universidad Nacional Micaela Bastidas de Apurímac. La UNAMBA, maneja un sistema transaccional el cual manipula datos operacionales los que por sí solos no generan información significativa para los directores, decanos, jefes de departamentos académicos ya que tienen que analizar la información para poder plantear planes de acción y estrategias de mejoras. Se propuso una solución de inteligencia de negocios para que las autoridades cuenten con una aplicación de gestión académica que proporcione información resumida, específica y precisa en tiempo real, que ayude a analizar, tomar decisiones y correcciones necesarias sobre el seguimiento y rendimiento académico de estudiantes, información de egresados, graduados y titulados, plana de docentes, y conocimiento de la población universitaria. Para el desarrollo de la solución de inteligencia de negocios se utilizó la herramienta de Pentaho y se trabajó con la metodología Hefesto, la cual está compuesta por los siguientes pasos: análisis de requerimientos, análisis de data sources, modelo lógico e integración de datos”.

Alexander (2016), en su tesis de titulación: “Elaboracion de un data mart para evidenciar el retraso academico en los alumnos de pregrado de la fii - unp” realiza:

“Una investigación que consistió en evidenciar el retraso académico que se .ha dado en los alumnos de la facultad de Ingeniería Industrial, desde la promoción 2006 hasta la 2014, cuyas escuelas pertenecientes son: Ingeniería Agroindustrial e Industrias Alimentarias, Ingeniería Industrial, Ingeniería Informática e Ingeniería Mecatrónica. Se aplicó la Tecnología de Business Intelligence, el cual se consideró necesario, para la creación del Data Mart específico al retraso académico. De la información que se tomó para la creación del Data Mart se transformó en conocimiento y así se logró tener una visión clara con respecto al retraso académico en la F

acuitad de Ingeniería Industrial. Con la implementación del Data Mart se pudo detallar dinámicamente el avance o el retraso académico del alumno, siendo este un apoyo al personal académico que lo solicite para la toma de decisiones hacia un pedido de beneficio del alumno, un informe de avance estudiantil por promociones, un informe completo por escuela y sus respectivas áreas, etc. La información evaluada pertenece a una cantidad de 114355 cursos inscritos en total, de los cuales se tiene como resultado los cursos Aprobados, Desaprobados y Retirados. Los resultados nos mostraron que la escuela que tiene menos cursos aprobados es Mecatrónica con un 70.33%, la escuela que tiene más cursos desaprobados es Informática con un 11.1 O% y la escuela que tiene más retirados es la de Mecatrónica con un 20.31”.

2.2 MARCO CONCEPTUAL

2.3 TOMA DE DESICIONES

"La organización es un sistema de decisiones en donde la gente participa conscientemente y racionalmente, escogiendo y decidiendo entre alternativas más o menos racionales que le son presentadas. Los directivos y gerentes de una empresa son los responsables de los hombres, dinero, maquinaria, materiales y métodos a su disposición, es por ello que para determinar su uso y alcanzar los objetivos de la empresa, deben de tomarse constantes decisiones que, en un momento dado, pueden tener repercusiones tanto internas (en cuanto a las utilidades, el producto, persona, etc.) como externas (relación con proveedores, la economía, el entorno, clientes, etc.) De la organización. La toma de decisiones es el proceso sistemático y racional a través del cual se selecciona entre varias alternativas el curso de acción más óptimo". (Carreto, 2010)

2.3.1 ETAPAS EN LA TOMA DE TOMA DESICIONES

El proceso de toma de decisiones conta de 4 etapas:

- Inteligencia: Consiste en realizar un análisis para su correcta identificación y comprensión del problema.
- Diseño: Consiste en elaborar en una serie de opciones como alternativas de solución

- Selección: Consiste en elegir la opción más adecuada de las alternativas de solución.
- Implementación: Consiste en gestionar la ejecución de la solución seleccionada y medir su impacto y posibles mejoras.

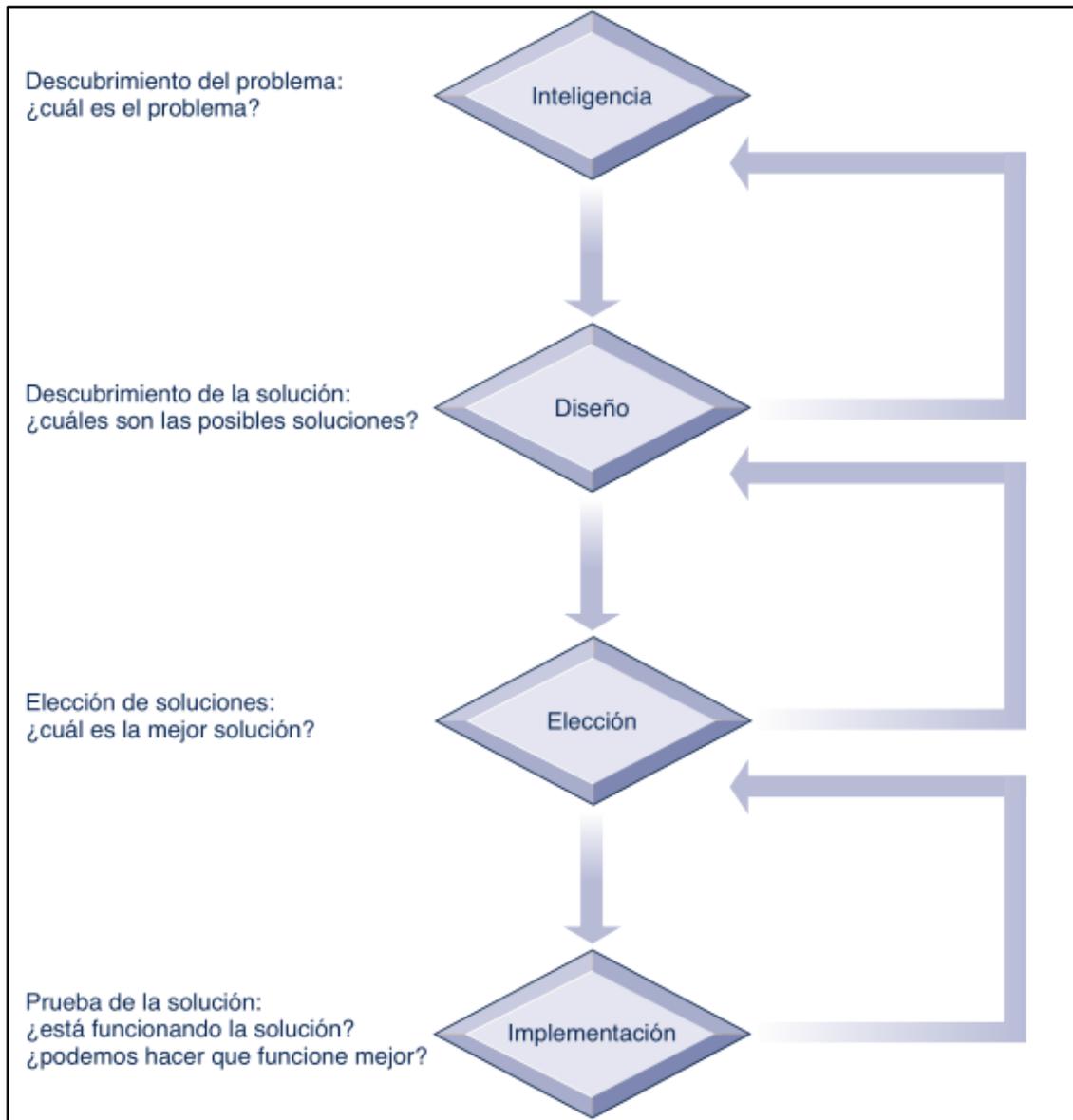


Figura 5: Etapas en la Toma de Decisiones

2.3.2 TIPOS DE TOMA DE DECISIONES

Las decisiones se clasifican como estructuradas, semiestructuradas y no estructuradas.

Las decisiones **no estructuradas** son aquellas en las que el responsable de seleccionar un decisión, debe realizar una evaluación de acuerdo a los requerimientos del negocios para resolver el problema. Cada una de estas decisiones no es frecuente y es no rutinaria, por lo que no hay un procedimiento bien establecido y flujo para tomarlas para tomarlas. En contraste, las **decisiones estructuradas** son frecuentes y rutinarias; además se requiere un procedimiento y flujo definido de tal forma que cada vez que haya que tomarlas, no se consideren como si fueran nuevas. Muchas decisiones tienen elementos de ambos tipos de decisiones y son **semiestructuradas**, en donde sólo una parte del problema tiene una respuesta clara proporcionada por un procedimiento aceptado. En general, las decisiones estructuradas son más frecuentes en los niveles más bajos de la organización, en tanto que los problemas no estructurados son más comunes en los niveles más altos de la organización.

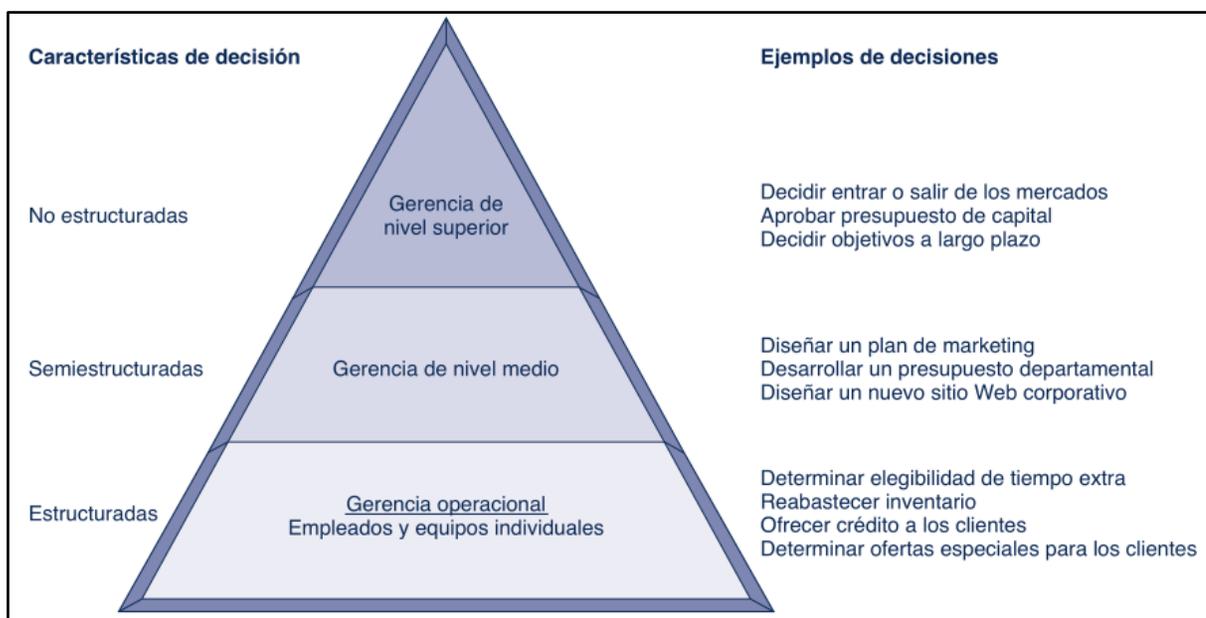


Figura 6: Tipos de Decisiones

2.4 INTELIGENCIA DE NEGOCIOS

Hoy en día la mayoría de las organizaciones cuentan con un volumen grande de datos, por ejemplo, de empleados, clientes, departamentos, ventas, compras, promociones, finanzas, entre muchos otros más. Si bien es cierto que estos datos provienen de diferentes orígenes, la mayoría de ellos son el resultado de las operaciones que se realizan a diario, es decir de los sistemas transaccionales con los que cuenta la organización. Sin embargo, en muchas ocasiones estos datos solamente se mantienen almacenados sin hacer un uso adecuado de ellos y peor aún sin sacar una ventaja competitiva. A diferencia de años atrás y gracias al uso del Internet y de las Tecnologías de la Información y Comunicaciones con las que cuentan las organizaciones actualmente, los directivos ahora pueden acceder a mucha más información de manera rápida, sencilla y mejor entendible. Las organizaciones de ahora buscan transformar los datos que tienen almacenados en información, y esta información a su vez en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones.

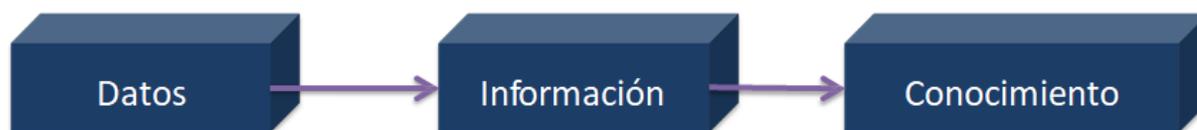


Figura 7: La Inteligencia de Negocios

“La Inteligencia de Negocios es un concepto que en los últimos años ha tomado gran importancia entre las organizaciones y desde luego dentro del área de la Informática, también es conocida como Business Intelligence por su término anglosajón y fue acuñado por un consultor de Gartner llamado Howard Dresner, el cual la definió como un proceso interactivo para explorar y analizar información estructurada sobre un área normalmente almacenada en un Data Warehouse, para descubrir tendencias o patrones, a partir de los cuales se pueden derivar ideas y extraer conclusiones” (Cano, 2007).

“The Data Warehouse Institute también propone una definición más amplia sobre Inteligencia de Negocios como un término que abarca los procesos, las herramientas y las tecnologías para convertir datos en información, la información en conocimiento y planes para

conducir de forma eficaz las actividades de los negocios. Abarca las tecnologías de Datawarehousing los procesos en el “back end”, consultas, informes, análisis y las herramientas para mostrar información y los procesos en el front end” (Oracle, 2016).

“Otra definición más pragmática de Inteligencia de Negocios que se tiene es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio” (Sinnexus, s.f.).

Actualmente la Inteligencia de Negocios es un factor estratégico para cualquier tipo de organización, ya que permite tener una ventaja que permite proporcionar una información relevante para responder de manera rápida los problemas del negocio.

2.5 SISTEMAS DE SOPORTE DE TOMA DE DECISIONES

“Un Sistema de Soporte a la Decisión (Decision Support System, DSS), es un sistema informático que utiliza información y modelos matemáticos para ayudar a tomar decisiones empresariales adecuadas según las condiciones del mercado y la situación interna de la organización” (Businessintelligence, 2013).

“Los DDS son sistemas de apoyo enfocados en el análisis de los datos de la organización y sirven para ayudar a los directivos a tomar decisiones que a menudo son cambiantes y se deben tomar rápidamente” (Mos, 2015).

Continuamente dentro de las organizaciones se toman decisiones de distinto tipo y de manera diferente, hay decisiones que se toman de manera informal casi por instinto que no afectan a la organización, sin embargo existen otras decisiones que se deben tomar de manera racional y absolutamente informada ya que pueden llegar a comprometer el funcionamiento y éxito de la

organización, por lo que el uso de un DSS ayuda a que se tengan resultados más eficientes y efectivos.

Los DSS tienen como principal objetivo explotar al máximo la información que se tiene almacenada dentro de una base de datos corporativa, ya sea un Data Warehouse o un Datamart, apoyándose de informes dinámicos y navegables a través de una interfaz gráfica, amigable y sencilla. Otra característica importante de los DSS es que pueden ser utilizados por usuarios de cualquier nivel gerencial dentro de una organización no solo por la alta dirección, tanto para situaciones estructuradas como no estructuradas.

2.5.1 SISTEMAS OLTP

“Un sistema OLTP (OnLine Transaction Processing) se puede definir como una herramienta tecnológica capaz de soportar el procesamiento, administración y mantenimiento diario de transacciones generadas por los negocios de una organización, para ofrecer altos niveles de disponibilidad, seguridad y confiabilidad. El proceso transaccional es típico de las bases de datos operacionales. Una transacción genera un proceso atómico que debe ser validado con un commit, o invalidado con un rollback, y que puede involucrar operaciones de inserción, modificación y borrado de datos. En las base de datos OLTP se encuentra la información actual y detallada del negocio, esta bases de datos por lo regular se implementan siguiendo el modelo relacional llevado a la 3FN” (Business intelligence, 2013).

2.5.2 SISTEMAS OLAP

“Un sistema OLAP (OnLine Analytical Processing) es una solución utilizada en el campo de la Inteligencia de Negocios cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales o cubos OLAP que contienen datos resumidos de grandes bases de datos o sistemas transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares” (Businessintelligence, 2013).

“En las bases de datos OLAP se encuentra información agregada e histórica que se almacena en esquemas multidimensionales como son el esquema estrella o copo de nieve. Se caracterizan por un volumen relativamente bajo de transacciones y las consultas frecuentemente son muy complejas e incluyen funciones de agregación” (Datawarehouse4u, 2009).

2.5.3 DATAWHAREHOUSE

Un Data Warehouse es un almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información. Los datos de un Data Warehouse deben almacenarse de forma segura, fiable, fácil de recuperar y de administrar.

El concepto de Data Warehouse se originó en 1988 con el trabajo de los investigadores de IBM, Barry Devlin y Paul Murphy aunque el término Data Warehouse fue acuñado por William H. Inmon, el cual es conocido como el padre de Data Warehousing. Inmon describió un Data Warehouse como una colección de datos orientada a un tema específico, integrado, variante en el tiempo y no volátil, que soporta el proceso de toma de decisiones.

“Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas” (Grupo PowerData, 2017).

“Proporciona una visión global, común e integrada de los datos de la organización, independiente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios. Normalmente en el Data Warehouse habrá que guardar información histórica que cubra un amplio período de tiempo” (El Rincon del BI, 2010).

“Un Data Warehouse es un contenedor en el que se almacenan los datos procedentes de las distintas fuentes que puedan existir en una organización, quedando éstos integrados, depurados y ordenados en una única base de datos centralizada. En este almacén se guardarán los datos durante el período de tiempo requerido para cumplir con las necesidades de consulta de cada organización. Con este sistema, las compañías consiguen tener integrados en un único

contenedor todos los datos de sus diferentes procesos de negocio, listos para ser analizados mediante las herramientas de explotación y reporting” (BI Geek, 2016).

2.5.4 DATAMART

“Un Data Mart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un Data Mart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información (Sinnexus, s.f.). Se ocupa de almacenar información de un departamento o grupo de trabajo específico. Funciona como una aplicación del Data Warehouse o una alternativa para empresas medianas que no pueden afrontar los costos de implementar un sistema tan amplio de almacenamiento de data. Las Data Marts pueden ser dependientes o independientes del Data Warehouse. Sin embargo, cabe mencionar que contar con sistemas independientes que no se encuentren integrados entre sí puede dificultar las tareas de administración y mantenimiento” (Business, 2015).

“El Data Mart, cuya definición es bastante similar a la del Data Warehouse, siendo su alcance la principal diferencia entre estos dos tipos de bases de datos. Así, mientras un Data Warehouse contiene todos los datos de una organización, un Data Mart solamente recoge un subconjunto de éstos, centrándose en un área específica dentro del negocio. Su objetivo es cubrir las necesidades de un determinado departamento dentro de la organización, por lo que podría definirse como un almacén de datos departamental. El Data Mart es un sistema orientado a la consulta, cuya distribución interna de los datos es clara y no hay dudas al respecto, estando éstos estructurados en modelos dimensionales de estrella o copo de nieve” (BI Geek, 2016).

2.5.5 METODOLOGÍAS DE INTELIGENCIA DE NEGOCIOS

2.5.5.1 METODOLOGÍA EFESTO

Es una metodología para la implementación de un data warehouse y esta compuesto de 4 fases con sus respectivas actividades:

- Análisis de requerimientos
- Análisis de los OLTP
- Modelo lógico del data warehouse
- Integración de datos

“Puede ser implementada con cualquier ciclo de vida cuyas fases de requerimientos y análisis no sean muy extensas para su implementación y cumpla con las necesidades de los usuarios” (Bernabeu, 2010). La Figura 6 lista las actividades dentro de cada fase de la metodología HEFESTO

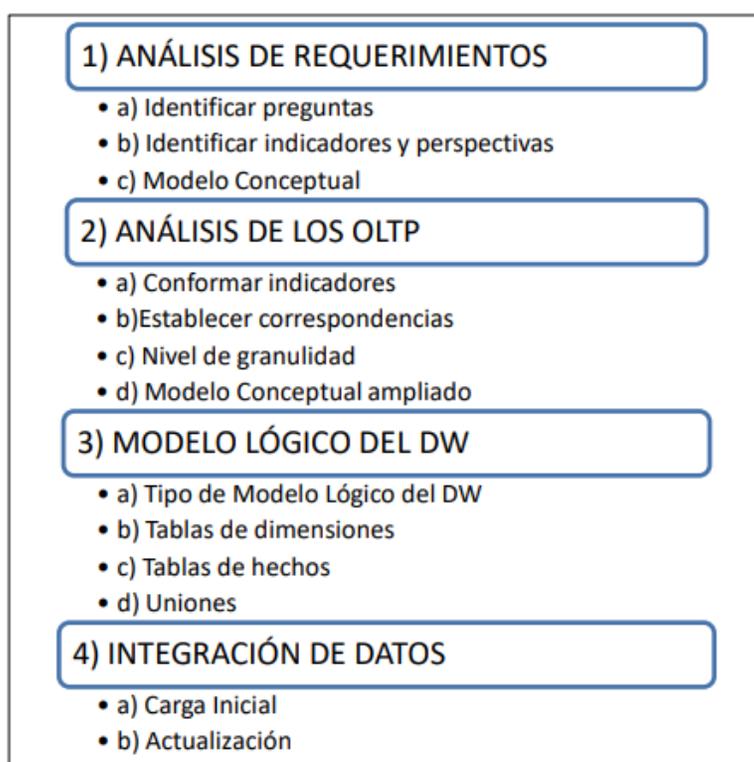


Figura 8: Actividades de la Metodología Hefesto

2.5.5.2 METODOLOGÍA DE BILL INMON

Bill Inmon tomó como punto esencial la transferencia de la información de los diferentes Sistemas Transaccionales de las organizaciones a un lugar donde todo está centralizado y los datos pueden ser utilizados para su estudio y análisis y que debe tener las siguientes características:

Orientado a temas. - Los datos deben estar interrelacionados de tal manera que sean una representación de los objetos del mundo real.

Integrado. - La base de datos contiene toda la información de la organización de manera integral y relacionada

No volátil. - La información debe tener la propiedad de ser consultada independiente del tiempo sin tener lagunas de información, debiendo ser accedida para futuros requerimientos de información.

Variante en el tiempo. - Los cambios deben registrarse con el fin de realizar el análisis dinámico de la información

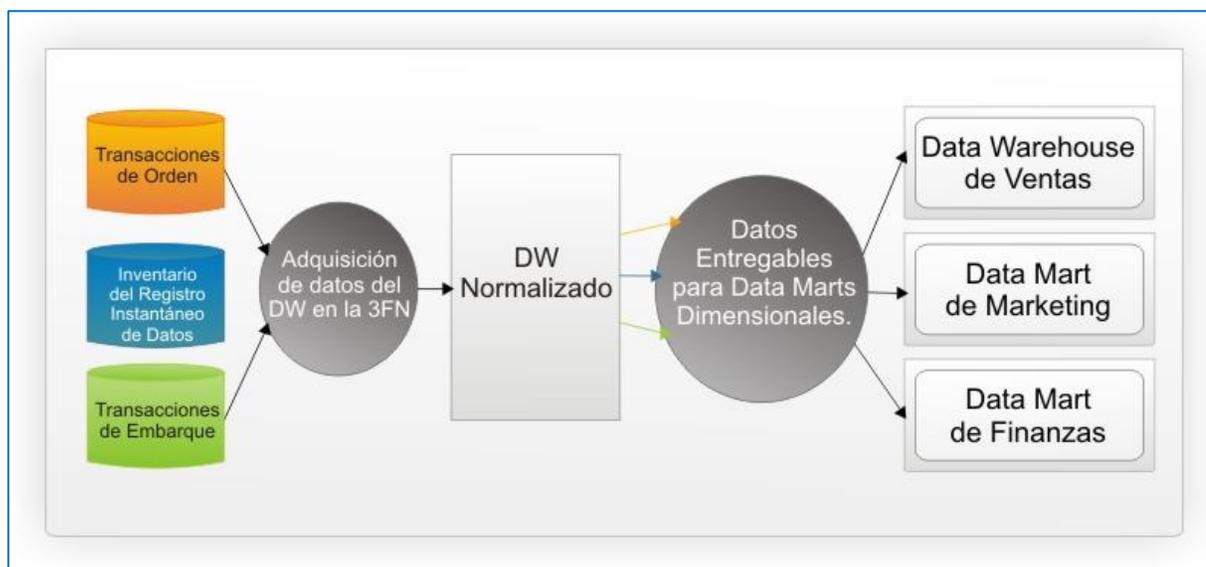


Figura 9: Dataware House Cooperativo de Bill Inmon

2.5.5.3 METODOLOGÍA DE RALPH KIMBALL

Ralph Kimball (1944) es considerado el inventor del Modelo Dimensional y pionero en Data Warehouse y la Inteligencia de Negocios. “Define un almacén de datos como: una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis. También fue Kimball quien determinó que un Data Warehouse no era más que: la unión de todos los Data Marts de una entidad”. “Defiende por tanto una metodología ascendente (bottom-up) a la hora

de diseñar un almacén de datos” (Brito, 2014). La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

- **“Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores”. (Kimball, 2011)
- **“Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa”. (Kimball, 2011)
- **“Realizar entregas en incrementos significativos:** Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software”. (Kimball, 2011)
- **“Ofrecer la solución completa:** Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios”. (Kimball, 2011)

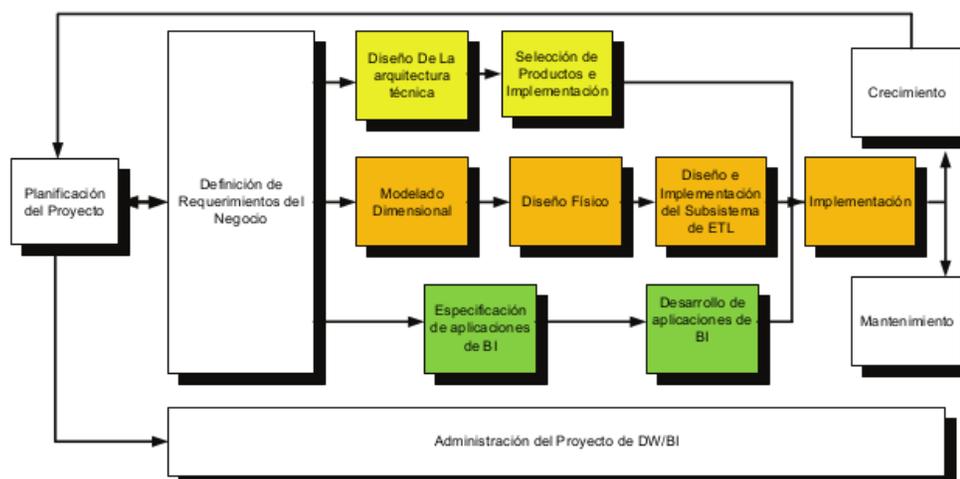


Figura 10: Metodología de Ralph Kimball

2.5.6 CUADRO COMPARATIVO DE METODOLOGIAS RELEVANTES

Tabla 2:
Tabla comparativa Mball vs Inmon

	Kimball	Inmon	Efesto
Diseño	Bottom-Up	Top-Down	Se adapta al Botton-UP y Top - Down
Flexibilidad	Media	Alto	Alto
Enfoque	Tiene un enfoque por procesos que son manejados por las diferentes áreas del proceso. Trata de responder necesidades específicas.	Tiene un enfoque global de toda la organización. No está basado en requerimientos específicos.	Se adapta por áreas y en el enfoque global
Costos	Implementar el Data Mart no presenta alto costo	Los costos son elevados ya implica trabajo de grandes cantidades de datos.	Costo inicial alto
Tiempo	Iterativo	Solo una vez	Solo una vez
Alcance	Departamentos individuales.	Toda la organización.	
Modelado	Estrella	Relacional	Estrella/Copo de Nieve

Fuente: Propia

2.5.7 PENTAHO

“Pentaho es una aplicación de software para la gestión de la inteligencia empresarial (Business Intelligence y Big Data). Está desarrollada con la filosofía “opensource” por lo que no tiene coste de licencias. Como plataforma cubre y satisface todos los requisitos de BI en términos de análisis y de gestión de datos, administración y seguridad. Ofrece soluciones para informes, análisis multidimensional (OLAP), minería de datos (Data Mining), tableros de

mando (Dashboard) y consultas ad-hoc. Cuenta con herramientas para la extracción de datos, transformación y carga (ETL)” (Itop.es, 2017). Pentaho, perteneciente a la empresa Hitachi Data Systems, es una empresa líder en integración de datos y análisis de negocio con una plataforma de clase empresarial, basado en open source. La misión de Pentaho es ayudar a las organizaciones de diferentes sectores a aprovechar el valor de todos sus datos, incluyendo Big Data e Internet of Things. Pentaho cuenta con más de 15.000 implementaciones de productos y 1.500 clientes comerciales actuales incluyendo EMC, Moody, o NASDAQ. La solución ha sido desarrollada en el lenguaje de programación Java y tiene una interfaz de implementación también basado en Java, que conlleva que Pentaho sea una solución muy usable al alcanzar una alta gama de necesidades de negocios empresariales.



Figura 11: Logo del Software Pentaho

2.5.7.1 FUNCIONALIDADES

Pentaho es una suite muy completa que cubre multitud de áreas analíticas y motores para el procesamiento y la obtención de conocimiento. Estas herramientas están soportadas e integradas sobre un servidor web y dispone de un entorno de configuración y desarrollo.

La suite Pentaho está compuesta por múltiples aplicaciones independientes entre sí, que trabajan juntas para crear y distribuir soluciones BI. Los principales proyectos que componen la suite son:

- Pentaho Business Analytics (Pentaho BA): es la plataforma web a la que se accede para ejecutar los recursos (informes, cuadros de mandos, etc.).
- Pentaho Report Designer (PRD): herramienta para crear reportes avanzados de alta calidad, interactivos y dinámicos.
- Pentaho Data Integration (PDI): herramienta de Integración de Datos. • Pentaho Analysis (Mondrian OLAP server): motor multidimensional utilizado para responder a las consultas de los OLAP.

- Pentaho Data Mining (Weka): herramienta de Data Mining.
- CTools: conjunto de herramientas para crear y gestionar Dashboards.
- Saiku: atractivo visor OLAP con una UI muy amena.

2.5.7.2 VENTAJAS

- Solución gratuita. Es una solución opensource. También dispone de una versión empresarial comercial.
- Completa. Incorpora todas aquellas herramientas necesarias para el BI, incluyendo Big Data e Internet of Things.
- Escalable. Plataforma robusta y con tecnología escalable.
- Flexible. Es posible elegir entre varias opciones dentro de cada área analítica.
- Comunidad. Amplia red de colaboradores.
- Internacional. Disponible en diversos idiomas.

2.5.7.3 PORQUE ELEGIR PENTAHO

- Líder mundial en software de BI opensource.
- Aparece en el Cuadrante Mágico de Gartner.
- Gran comunidad de seguidores.
- Nos permite optimizar costos de implementación

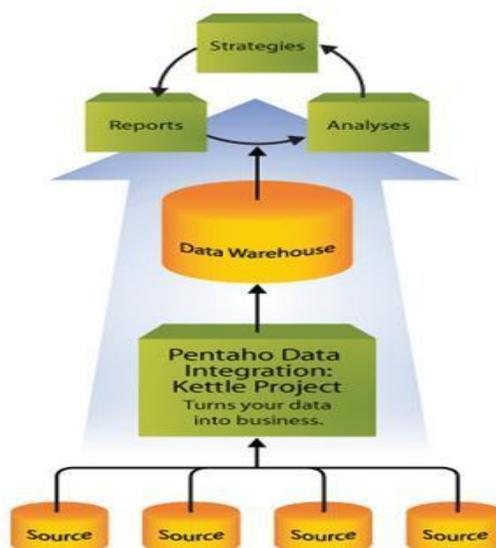


Figura 12: Arquitectura de Pentaho

2.5.8 MINERIA DE DATOS

No existe una única definición del concepto de minería de datos, aquí algunas posiciones al respecto.

- “MD es la extracción de información implícita, desconocida o previamente ignorada, que puede ser útil de un conjunto de datos” (Vilches González, et al, 2007).
- MD lies at the interface of statics, database, pattern recognition and machine learning” (Riganello, et al, 2010).
- “El proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos” (Hernández Orallo, 2007)

Se puede resumir como las técnicas que permiten ver más allá que simples datos si no más bien reconocimiento de patrones de información que sean útil para los interesados. De manera análogo como los hace la minería en la extracción de los minerales de la tierra. Por ello es necesario una serie de técnicas y modelos.

2.5.8.1 AREA DE APLICACIÓN

Marketing:

- Para identificar patrones de compra y la segmentación de clientes.
- Predecir la respuesta de las campañas de mailing.

Compañías de Seguros:

- Para determinar los procedimientos médicos solicitados.
- Predecir clientes que compran polizas.
- Identificar comportamiento fraudulento

Banca

- Patrones de uso fraudulento de tarjetas de crédito
- Identificar clientes leales.
- Determinar gastos de en tarjetas de crédito.
- Determinar correlaciones entre índices financieros.

Telecomunicaciones

- Detección de fraude telefónico.

Medicina

- Identificación de terapias médicas satisfactorias para diferentes enfermedades
- Asociación de síntomas y clasificación de patologías
- Estudios de factores de riesgos para la salud

Industria farmacéutica

- Clustering de moléculas.
- Búsqueda de patrón de moléculas
- Predicción de resultados de experimentos

Biología

- Con la decodificación del Proyecto GNOMA, la siguiente tarea consiste en entender cómo funcionan nuestros genes. Existen nuevas tecnologías (chips de ADN, proteómica, genómica funcional, variabilidad genética individual) que están promoviendo el desarrollo de una “nueva biología” que permite extraer conocimientos biomédicos a partir de bases de datos experimentales en el entorno de un ordenador básicamente mediante técnicas de minería de datos y visualización. Estos trabajos forman parte de los desarrollos de la Bioinformática.

2.5.8.2 TIPOLOGÍA DE ALGORITMOS

Los algoritmos de minería de datos se dividen en dos grandes grupos:

- Los **métodos supervisados**, que requieren de un conjunto de datos previamente etiquetados con el conjunto de clases
- Los **métodos no supervisados**, donde los datos no tiene ninguna etiqueta o clasificación previamente.

En la siguiente figura se muestra los principales métodos de minería de datos indicando en cada caso la tipología y de la tarea que puede resolver

2.5.8.2.1 METODOS NO SUPERVISADOS

- a. **Clustering o agrupamiento:** Es el proceso para agrupar datos de un dominio dado, descrito por un número de atributos discretos o de valor continuo, en un conjunto de grupos (clústeres) basándose en la similitud entre las instancias. La función de agrupamiento o clustering se puede modelar mediante

$$h: X \rightarrow C_h$$

Donde h representa la función de agrupamiento, X el conjunto de atributos que forman una instancia y C_h un conjunto de grupos clústeres.

- b. **Reglas de asociación:** Es la exploración de datos con el propósito de identificar relaciones entre los datos, dentro de una fuente o base de datos. Son utilizados para análisis explorativos, buscando relaciones dentro del conjunto.

2.5.8.2.2 PRINCIPALES TECNICAS SUPERVISADAS

- a. **Regresión:** Es una tarea de aprendizaje inductivo que permite predecir valores numéricos en lugar de clases discretas
- b. **Arboles de decisión:** Representan estructuras conjuntas de decisiones y estas decisiones generan reglas de clasificación.
- c. **Clasificación:** Técnica que permite encontrar modelos que describen y distinguen conjunto de clases o conceptos para futuras predicciones
- d. **Redes Neuronales:** Las redes neuronales simulan el cerebro humano mediante el aprendizaje de un conjunto de datos de entrenamiento

Tabla 3: Tipología de los algoritmos

Métodos	Supervisados		No supervisados
	Clasificación	Regresión	Agrupamiento
Agrupamiento jerárquico			X
k-means y derivados			X
k-NN	X		
SVN	X	X	
Redes Neuronales	X	X	
Arboles de decisión	X	X	
Métodos probabilísticos	X	X	

2.5.8.3 ENTRENAMIENTO Y TEST

Para validar un algoritmo de aprendizaje o modelo es necesario asegurar que funcionará correctamente para los datos de prueba o test futuros, de forma que capture la esencia del programa a resolver y generalice correctamente. Es decir que se trata de evitar que sea dependiente de los datos utilizados en su entrenamiento evitando el problema de sobre entrenamiento (en inglés, *overfitting*). Cuando hablamos de datos de entrenamiento y de test, estamos refiriéndonos exclusiva a los algoritmos de aprendizaje supervisado, ya que es necesario evaluar los resultados obtenidos sobre datos etiquetados nunca vistos anteriormente y se pueden comparar los errores cometidos.

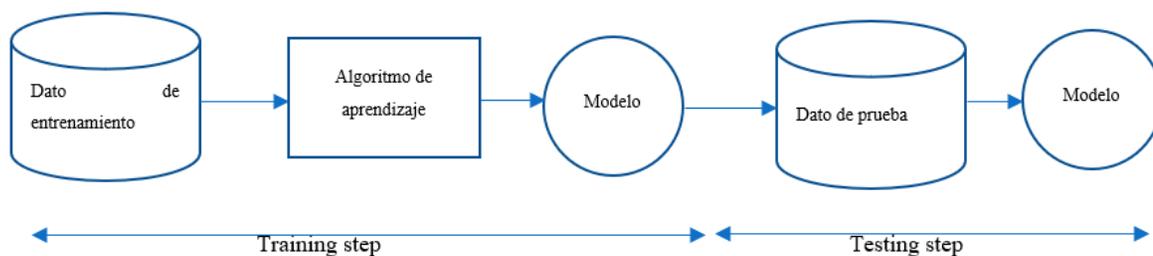


Figura 13: Proceso de creación y validación de un modelo basado en aprendizaje supervisado

Leave-one-out: Tal y como su nombre indica, la técnica leave-one-out consiste en utilizar todos los datos menos uno para construir el modelo y utilizar el dato no usado para evaluarlo. Obviamente, si el número de datos es n , este proceso puede repetirse n veces, siendo posible calcular el error promedio de los modelos construidos

Leave-p-out: Una generalización del método anterior es utilizar un subconjunto de $p < n$, de forma que se utilizan $n-p$ datos para entrenar el modelo y p para validarlo, siguiendo el mismo proceso. El problema es que el número de subconjuntos posibles de p elementos tomados de un conjunto de n crece exponencialmente, con lo cual para $n/$ y/o p grandes el problema puede ser intratable computacionalmente, por lo que p suele ser muy pequeño.

k-fold cross validation: Para conjuntos del orden de cientos de miles o miles de muestras, una solución de compromiso es realizar una partición aleatoria del conjunto de datos en k conjuntos del mismo tamaño, usando $k-1$ conjuntos para entrenar el modelo y el conjunto restante para evaluarlo, repitiendo el proceso k veces y promediando el error estimado. Habitualmente, se escoge un valor de $k=5$ o $k=10$, aunque no existe ninguna base teórica que sustente dichos valores, si no que es el resultado de la experimentación.

2.5.8.4 PRINCIPALES TECNICAS DE MINERIA DE DATOS DE CLASIFICACIÓN

2.5.8.4.1 REDES NEURONALES ARTIFICIALES

(García Serrano, 2014) nos dice “Una neurona es una célula del cerebro cuya función es la recepción, procesamiento y emisión de señales eléctricas. Una neurona está compuesta por el núcleo, que no es más que una célula especializada, rodeada de millones de conexiones que la unen a otras neuronas. Estas conexiones se denominan sinapsis. Cada neurona está conectada de media a otras mil, aunque pueden ser muchas más. Se estima que un niño tiene alrededor de 1,000 billones de conexiones que con el tiempo se van perdiendo hasta alcanzar alrededor de los 500 billones en un adulto. Las conexiones se realizan mediante dos tipos de neurotransmisores: las dendritas y los axones. Según la neurociencia actual, parece que una

neurona funciona de forma muy similar a un transistor, es decir en un momento puede estar activa o no activa. Las dendritas transportan señales eléctricas desde otras neuronas. Cuando la cantidad de dendritas que envían señales alcanzan un umbral determinado, la neurona se activa y envía señales eléctricas a otras neuronas a través de los axones. Es decir, funcionan como una autentica red con billones de conexiones que trabajan de forma paralela”.

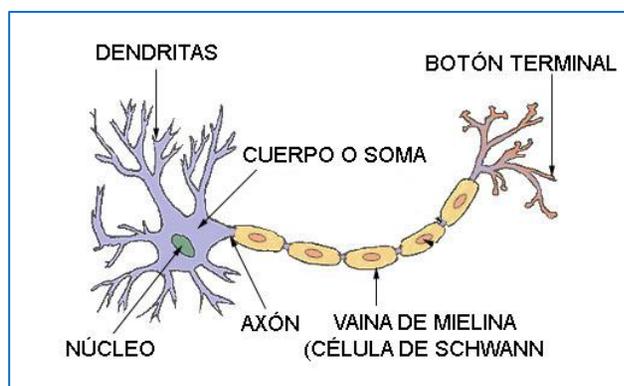


Figura 14: Partes de una red neuronal biológica

A continuación, se representa un modelo matemático de una red neuronal

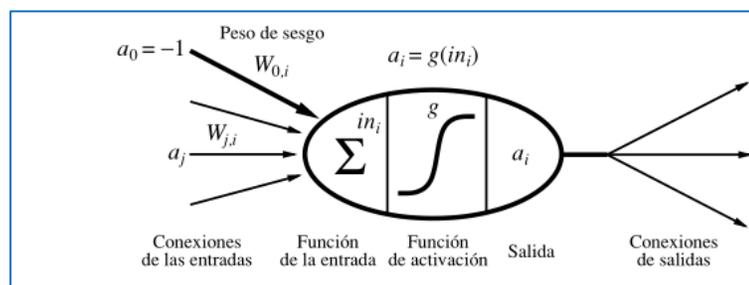


Figura 15: Modelo matemático de una Red Neuronal Artificial sencilla

A. REDES NEURONALES DE UNA SOLA CAPA (PERCEPTRONES)

Una red con todas las entradas conectadas directamente a las salidas se denomina “red neuronal de una sola capa o perceptrón”, como se muestra a continuación.

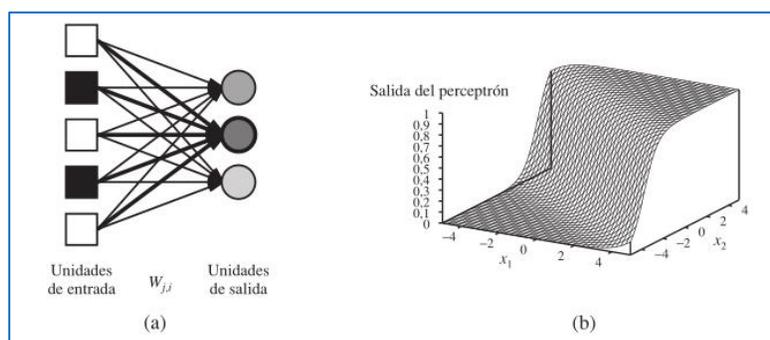


Figura 16: (a) Red perceptrón con 5 entradas y 3 salidas. (b) Un grafico de salida del perceptrón de 2 unidades

B. REDES NEURONALES MULTICAPA CON ALIMENTACION HACIA ADELANTE

En este caso se considera redes con unidades ocultas, el cual da la ventaja es que se amplía el espacio de hipótesis que se pueden representar con la red neuronal de una sola capa.

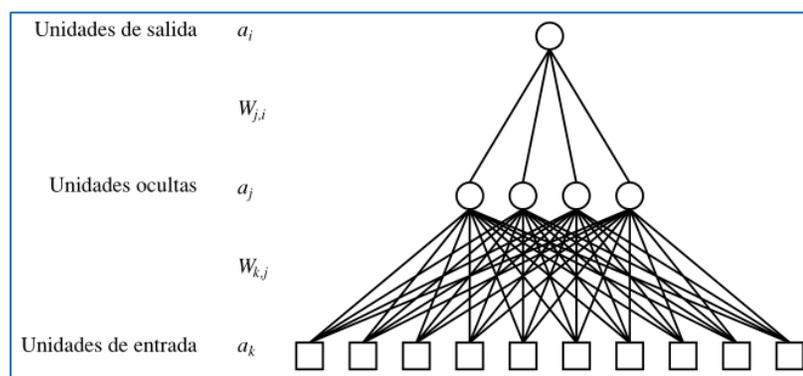


Figura 17: Red Neuronal Artificial con una capa oculta y 10 entradas

2.5.8.4.2 ARBOLES DE DECISION

Los árboles de decisión es uno de los algoritmos más sencillos y sencillos y con más éxito para construir algoritmos de aprendizaje por su alta capacidad explicativa y la facilidad para interpretar el modelo generado. Un árbol de decisión toma como flujo de entrada un conjunto de atributos y devuelve una “decisión”, el valor previsto de la salida dada la entrada. Un árbol de decisión

desarrolla una secuencia de test para poder alcanzar una decisión adecuada. Cada nodo interno del árbol corresponde con un test sobre el valor de una de las propiedades, y las ramas que salen del nodo están etiquetadas con los posibles valores de dicha propiedad. Cada nodo hoja del árbol representa el valor que ha de ser devuelto si dicho nodo hoja es alcanzado.

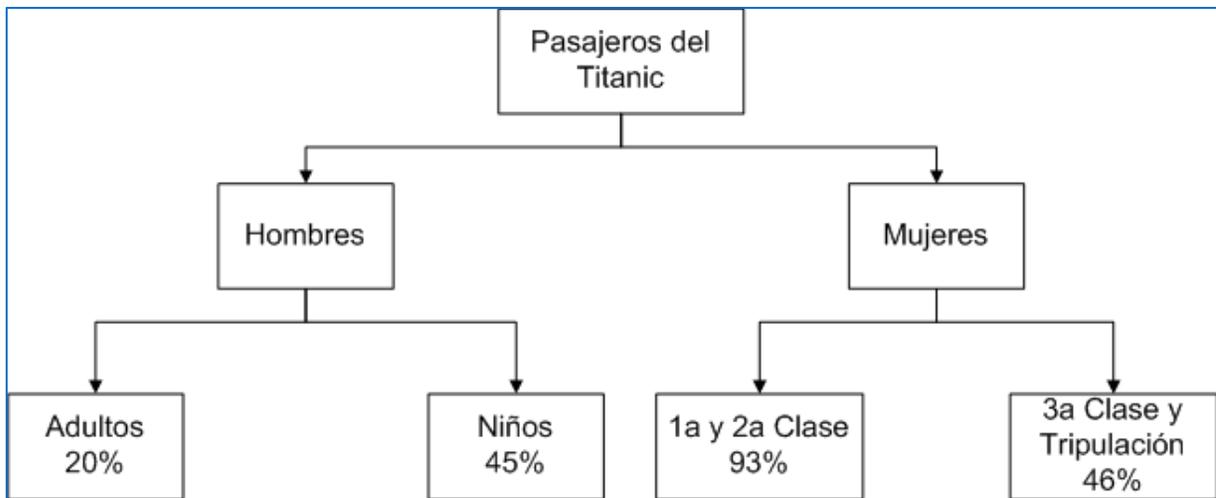


Figura 18: Supervivencia de los embarcados en el Titanic

A. CRITERIO DE SELECCIÓN:

Según (Girones, Casas Roma, Minguillon Alfonso, & Caihuelas Quile, 2017) indica que “El criterio de selección determina que nodo de hoja es escogido para ser particionado. Obviamente, una hoja que solamente contenga elementos de una misma clase no debe ser particionado, ya que no mejora la capacidad predictiva del árbol. El criterio es seleccionar el nodo más impuro, es decir aquel que contiene una mayor mezcla de elementos de diferentes clases”. Esto se mide con la entropía que mide el grado de desorden de una distribución de elementos de k clases diferentes:

$$H = \sum_{i=1}^k p_i \log p_i$$

La entropía es cero si todos los elementos son de una misma clase c

B. CRITERIO DE CLASIFICACION: El criterio de clasificación determina que clase se asigna a una región u hoja. Este criterio determina el error cometido en aquella hoja y también el error global que comete el árbol de decisión. Obviamente, si una hoja solamente contiene elementos de una clase, dicha clase es la elegida como representante de la región, minimizando el error cometido en la hoja, el cual es cero. En el caso de que en una hoja existan elementos diferentes clases, se escogerá aquella clave que minimiza el error cometido, normalmente la más poblada. En general, la clase escogida es aquella que satisface la siguiente ecuación:

$$c(t) = \underset{i}{\operatorname{argmin}} \sum_{i=1}^k p_i(t) C_{ij}$$

Es decir, se trata de escoger aquella clase que minimiza el error cometido, teniendo en cuenta el posible diferente de coste C_{ij} de cometer un error al escoger una clase i en lugar de la otra j , es decir con costes asimétricos ($C_{ij} \neq C_{ji}$)

Por ejemplo, en un árbol de decisión que intenta determinar la resonancia magnética de una muestra de tejido si esta se trata de un tumor o no, es mejor cometer y obtener falsos positivos (es decir, indicar que hay un tumor no es así) que no son falsos negativos (es decir, indicar que no se trata de un tumor cuando realmente sí lo es.)

El error cometido por un árbol T es la suma ponderada del error cometido en cada hoja, teniendo en cuenta el tamaño de la región definido por cada hoja, estimado a través del número de elementos que la componen.

$$e(t) = \frac{1}{n} \sum_t n_t \sum_{i=1}^k p_i(t) C_{ij}$$

2.5.9 RENDIMIENTO ACADEMICO

“El Rendimiento Académico se define como el producto de la asimilación del contenido de los programas de estudio, expresado en calificaciones dentro de una escala convencional. Así también el rendimiento académico relaciona la acción del proceso educativo, no sólo en el aspecto cognoscitivo logrado por el educando, sino también en el conjunto de habilidades, destrezas, aptitudes, etc. Con esta síntesis están los esfuerzos de la sociedad, del profesor y del proceso enseñanza-aprendizaje; el profesor es el responsable en gran parte del rendimiento escolar. Intervienen en este una serie de factores, entre ellos, la metodología del profesor, el aspecto individual del alumno, el apoyo familiar, la situación social, entre otros”. (Figuroa 2001).

Partiendo del punto de vista de Carlos Figuroa (2004, Pág. 25) que define el Rendimiento Académico como “el conjunto de transformaciones operadas en el educando, a través del proceso enseñanza-aprendizaje, que se manifiesta mediante el crecimiento y enriquecimiento de la personalidad en formación”.

2.6 ASPECTOS DE RESPONSABILIDAD SOCIAL Y MEDIO AMBIENTE

La implementación de la solución no afectará al equilibrio del medio ambiente dado que la interacción que tiene el software no afectará ningún aspecto del medio ambiente, toda vez que las salidas de la implementación serán a través de una plataforma digital.

3. CAPITULO III: METODO

3.1 TIPO Y NIVEL DE INVESTIGACIÓN

3.1.1 TIPO DE INVESTIGACIÓN

Aplicada: La investigación es aplicada porque aplica conocimientos de metodología de inteligencia de negocios en el proceso de Toma de Decisiones del Área Académica de la Universidad Nacional de Cañete.

3.1.2 NIVEL DE INVESTGACIÓN

La presente investigación es Explicativa y descriptivo.

3.1.3 DISEÑO DE INVESTIGACIÓN

Para dicho análisis se aplicará el diseño Pre-Experimental

G_e O₁ X O₂

Donde:

G_e = Grupo Experimental: Es el grupo de estudio al que se le asignará el estímulo (metodología de inteligencia de negocios).

O₁ = Datos de Pre-Prueba relacionados con los indicadores de la variable dependiente antes de implementarse la metodología de inteligencia de negocios.

O₂ = Datos de la Post-Prueba relacionados con los indicadores de la Variable Dependiente después de implementarse la metodología de inteligencia de negocios.

X = Metodología de inteligencia de negocios: Estímulo o condición experimental.

Descripción:

Debe medirse los indicadores de la variable dependiente (O₁) en el grupo experimental (G_e) constituido por el número de procesos de Toma de Decisiones antes de implementarse la solución de metodología de inteligencia de negocios en el proceso de toma de decisiones (X) para mejorarlo, luego una vez administrado el estilo se mide nuevamente los valores de cada

indicador de la variable dependiente (O_2). Por lo que se espera que los valores O_2 sean mejores que los valores de O_1 .

3.2 POBLACIÓN Y MUESTRA

3.2.1 POBLACIÓN

Se selecciona como unidades de análisis a los procesos de tomas de decisiones del Área
N= indeterminado

3.2.2 MUESTRA

Para esta investigación se realizará se ha tomado la decisión de tomar una muestra de 30 toma de decisiones ya que se trata de un valor referencial de acuerdo a la bibliografía consultada

$n = 30$ procesos de tomas de decisiones.¹

3.3 HIPOTESIS

3.3.1 HIPOTESIS GENERAL:

Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete.

3.3.2 HIPOTESIS ESPECÍFICAS:

A. Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del tiempo de generación de reportes del proceso de toma de decisiones del rendimiento académico.

¹ Pande, P. Las claves Prácticas de SIX SIGMA, 2004, pp. 135-136.

- B. Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en el tiempo de análisis de información del proceso de toma de decisiones del rendimiento académico.
- C. Si se aplica la metodología de metodología de inteligencia de negocios entonces influye positivamente en la mejora del nivel de satisfacción del usuario del proceso de toma de decisiones del rendimiento académico.

3.4 OPERACIONALIZACIÓN DE VARIABLES

3.4.1 VARIABLE INDEPENDIENTE:

Metodología de inteligencia de negocios

DEFINICIÓN CONCEPTUAL

La inteligencia de negocios es el conjunto de componentes que permiten optimizar la obtención rápida de conocimiento y que nos permitan tener respuestas para su análisis e interpretación, de manera que puedan ser aprovechados para la toma de decisiones y se conviertan en conocimiento para los tomadores de decisiones del negocio.

DEFINICIÓN OPERACIONAL

TABLA 4:
Operacionalización de Variable Independiente

Variable independiente	Indicadores	Índice
Metodología de inteligencia de negocios	Presencia - Ausencia	No, Si

Fuente: Elaboración Propia

3.4.2 VARIABLE DEPENDIENTE:

Proceso de toma de decisiones del rendimiento académico de la Universidad Nacional de Cañete.

DEFINICIÓN CONCEPTUAL

La toma de decisiones es un proceso establecido a través del cual se selecciona una alternativa de entre varias, siendo la seleccionada la óptima que nos permita llegar a nuestro propósito).

DEFINICIÓN OPERACIONAL

TABLA 5:
Operacionalización de Variable Dependiente.

Variable independiente	Indicadores	Unidad de Medida	Unidad de Observación
Proceso de toma de decisiones del rendimiento académico de la Universidad Nacional de Cañete	Tiempo de generación de reportes.	Segundos	Reloj
	Tiempo de análisis de la información	Minutos	Reloj
	Nivel de Satisfacción	----	Responsables académicos

Fuente: Elaboración Propia

3.5 INSTRUMENTOS

3.5.1 TÉCNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN DE CAMPO

TABLA 6:
Técnicas e Instrumentos de la Investigación de Campo.

Técnicas	Instrumentos
1. Observación Directa	Ficha de Registro de KPI'S
2. Aplicación de Cuestionarios	Cuestionario (documento)
3. Revisión de: Artículos Tesis Bibliotecas virtuales	Computador

Fuente: Elaboración Propia

3.5.2 TÉCNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN EXPERIMENTAL

TABLA 7:
Técnicas e instrumentos de la investigación experimental

Técnicas	Instrumentos
- Tiempo de generación de reportes.	Reloj
- Tiempo de análisis de la información	Reloj
- Satisfacción del uso del usuario	Cuestionario

Fuente: Elaboración Propia

3.5.3 TÉCNICAS E INSTRUMENTOS RELACIONADAS PARA LA INVESTIGACIÓN DOCUMENTAL

TABLA 8:
Técnicas e Instrumentos de la Investigación Documental.

Técnicas	Instrumentos
Revisión de:	Laptop y Smartphone
Tesis	USB
Libros	Fotocopias
Artículos Científicos	Impresiones
Monografía en Internet	Cuaderno de Apuntes

Fuente: Elaboración Propia

3.6 PROCEDIMIENTOS

BI_BM es una metodología de desarrollo de soluciones para la toma de decisiones el cual toma técnicas de Inteligencias de Negocio y Minería de datos para lo cual presentamos las etapas de desarrollo de la metodología BI_DM.

3.6.1 ESTUDIO DE PLANIFICACIÓN

3.6.1.1 FACTIBILIDAD TÉCNICA

La Factibilidad Técnica permite realizar un análisis de disponibilidad de software para la llevar a cabo el presente proyecto y con el objetivo de minimizar los costos de implementación determino las siguientes aplicaciones:

Motos de Base de Datos Mysql

Diseñador de Base de Datos: Mysql WorkBench

Pentaho Data Integración

Pentaho Server
Apache Tomcat
Weka

3.6.1.2 FACTIBILIDAD OPERATIVA

Actualmente se tiene es factible operativamente, ya que actualmente se cuenta con los conocimientos tecnológicos para el la implementación de la solución de inteligencia de negocios el cual tendrá una influencia positiva debido al requerimiento de información debido al proceso de Licenciamiento y Acreditación de las Carreras que se ha iniciado. Además de ello es factible por las siguientes razones:

La necesidad de la Vicepresidencia Académica en tomar decisiones de forma rápida y acertada

De acuerdo al proceso de recolección de información se realizaron entrevistas con los responsables de desarrollo, para analizar sus cuellos de botellas en el desarrollo de sus reportes.

La administración de la actual solución estará a cargo del Vicepresidente Académico.

3.6.1.3 FACTIBILIDAD ECONÓMICA

El presente trabajo de investigación es factiblemente económico debido a que a Universidad Nacional de Cañete se cuenta con el personal humano disponible y además de ello se está utilizando Open Source y software libre que disminuye el precio de las licencias de software.

3.6.2 PLANIFICACIÓN DEL PROYECTO

3.6.2.1 VISIÓN DEL PRODUCTO

Geográficos: El presente proyecto beneficiará al proceso de toma de decisiones del rendimiento académico de la Universidad Nacional de Cañete.

Organizacional: Optimizara los procesos de toma de decisiones en el Área de Vice presidencia académica para la toma de decisiones.

Funcional: Permitirá tener una herramienta útil para los requerimientos de información de forma eficaz.

Beneficios:

- Optimizara el proceso de toma de decisiones
- Aumentará la satisfacción de los usuarios
- Los tomadores de decisiones podrán utilizar la herramienta desde cualquier punto conectado a Internet, teniendo la información de manera oportuna
- Con la información oportuna se podrán programar acciones para mejorar el rendimiento académico de los alumnos de la Universidad Nacional de Cañete
- Mejorará la precisión de información

3.6.2.2 OBJETIVO DEL PROYECTO

Optimizar el tiempo en la generación de reportes relacionados al rendimiento académico. Brindar a la Vice Presidencia Académica una herramienta de fácil visualización del estado actual del rendimiento académico de forma sencilla que le permite programar actividades de mejora académica.

3.6.2.3 ALCALCE DELPROYECTO

- Desarrollo de un Data Mart para el área de la Vice Presidencia Académica de la Universidad Nacional de Cañete con los datos obtenidos del Sistema Transaccional de Gestión Académica.

- Elaboración de la base de datos dimensional.

- Diseño de reportes:

Reporte de asignaturas matriculadas por curso

Reporte de asignaturas aprobadas

Reporte de asignaturas desaprobadas

Reporte de créditos matriculados

Reporte de créditos aprobados

Reporte de créditos desaprobados

Reporte de promedio ponderado

3.6.2.4 EQUIPOS DE TRABAJO

A continuación, determinamos el grupo de trabajo formado por los siguientes roles:

Tabla 9:
Técnicas e Instrumentos de la Investigación Experimental.

Formación de equipo de trabajo	
ROL	DESCRIPCIÓN
Administrador del proyecto	Es el responsable de gestionar los tiempo, hitos, recursos y monitoreo constante del equipo
Analista de datos	Es la persona quien se encarga de recopilar y sistematizar la información obtenida de los usuarios para su implementación en la base de datos
Diseñador	Es la persona que diseña el procedimiento del ETL en relación a los requerimientos.
Implementadores	Grupo de desarrolladores que tiene la función e implementar la solución de BI

Fuente: Elaboración Propia

3.6.2.5 HARDWARE Y SOFTWARE

Hardware

Servidor Principal



Figura 19: Servidor Pro Eliant DL 3000

Tabla 10:

Especificaciones Técnicas del Servidor Pro Eliant DL 3000

Característica	Especificaciones técnicas
Número de procesadores	4 procesadores principales
Procesador Core	Intel Xeon® CPU E5-2403 0 @ 1.80Ghz
Memoria RAM	8 GB
Controlador de Red	Adaptador Ethernet NC382i multifunción de 1 Gb
Software de gestión	Microsoft Windows Server 2008 R2 Standard
Tipo de Unidad Optimca	No

Fuente: Elaboración Propia

Pcs de Usuario

Tabla 11:

Especificaciones Técnicas de Pcs de usuarios

Características	Especificaciones técnicas
Número de procesadores	1 procesador de 4 núcleos
Procesador Core	Intel Core(TM) i7-4790 CPU
Memoria	8.00 GB
Controlador de Red	Intel Ethernet Connection I217-V
Software de gestión	Windows 8.1 pro
Número de procesadores	1 procesador de 4 núcleos

Fuente: Elaboración Propia

Software

Tabla 12:
Lista de Software

Software	Funcionalidad
Mysql Server	Base de Datos donde se guarde los datos
Pentaho Server	Servidor OLAP
Pentaho Data Integration	Herramienta que permite realizar ETL
Pentaho Mondrian	Permite definir los Cubos de Información
Apache Tomcat	Contenedor de servlets desarrollado bajo el proyecto Jakarta en la Servidor de aplicaciones
Weka	Waikato Environment for Knowledge Analysis – Entorno para Análisis del Conocimiento de la Universidad de Waikato

Fuente: Elaboración Propia

3.6.2.6 ANÁLISIS DE RIESGOS

- Problemas de configuración con la herramienta de solución.
- Falta de integridad de datos de la Base de Datos Transaccional.

3.6.3 DEFINICIÓN DE REQUERIMIENTOS DEL NEGOCIOS DEL BI_DM

TABLA 13:
Lista de requerimientos

Código	Requerimiento
REQ 1	Visualizar las asignaturas matriculadas
REQ 2	Visualizar asignaturas aprobadas
REQ 3	Visualizar asignaturas desaprobadas
REQ 4	Visualizar créditos matriculados
REQ 5	Visualizar créditos aprobados
REQ 6	Visualizar créditos desaprobados
REQ 7	Visualizar promedio ponderado

Fuente: Elaboración Propia

3.6.4 DISEÑO DE LA ARQUITECTURA TECNICA DEL BI_DM

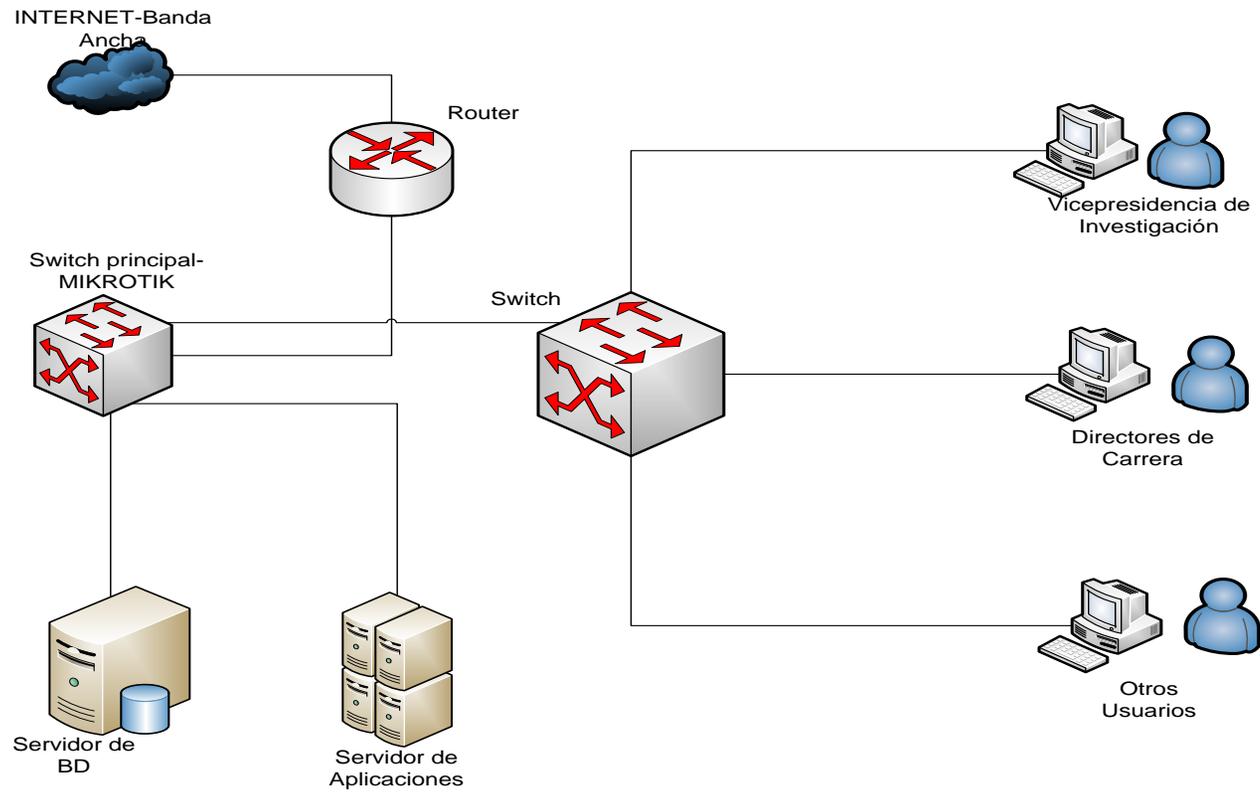


Figura 20: Diseño de la Arquitectura Técnica

3.6.5 SELECCIÓN DE PRODUCTOS E IMPLEMENTACIÓN

TABLA 14:
Rol de los Productos de Software

Nombre del Producto	Rol del producto
Mysql Server	Servidor de Base de Datos
Pentaho Server	Servidor OLAP
Pentaho Data Integration	Herramienta ETL
Pentaho Mondrian	Diseñador de Cubos
Apache Tomcat	Servidor de aplicaciones

Fuente: Elaboración propia

3.6.6 MODELADO DIMENSIONAL

3.6.6.1 MEDIDAS

Tabla 15:
Medidas y formulas

Medidas	Formulas
Asignaturas Matriculados por Curso	Count(v.id_ficha) as asignaturas_matriculadas,
Asignaturas aprobadas	Sum((case when v.promedio>10 then '1' when v.promedio<=10 then '0' end))
Asignaturas desaprobadas	Sum((case when v.promedio<=10 then '1' when v.promedio>10 then '0' end))
Créditos matriculados	Sum(v.creditos) as creditos_matriculados,
Créditos aprobados	Sum((case when v.promedio>10 then v.creditos when v.promedio<=10 then '0' end))
Créditos desaprobados	Sum((case when v.promedio<=10 then v.creditos when v.promedio>10 then '0' end))
Promedio Ponderado	Sum(v.creditos*v.promedio)/sum(v.creditos)

Fuente: Elaboración propia

3.6.6.2 DIMENSIONES

3.6.6.2.1 DIMENSIONES Y MEDIDAS

Tabla 16:
Dimensiones y medidas

MEDIDAS	DIMENSIONES				
	Dim_alumno	Dim_periodo	Dim_carrera	Dim_asignatura	Dim_ciclo
Asignaturas matriculadas por Curso	X	X	X	X	X
Asignaturas aprobadas	X	X	X	X	X
Asignaturas desaprobadas	X	X	X	X	X
Créditos matriculados	X	X	X	X	X
Créditos aprobados	X	X	X	X	X
Créditos desaprobados	X	X	X	X	X
Promedio Ponderado	X	X	X	X	X

Fuente: Elaboración propia

DIMENSIÓN ALUMNO

Tabla 17:
Dimensión Alumno

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Identificador	Identificación del alumno	34
Nombre	Nombre completo del contribuyente	Luyo Sanchez María
Departamento	Departamento de origen del alumno	Lima
Provincia	Provincia de origen del alumno	Cañete
Distrito	Distrito de origen del alumno	San Vicente

Fuente: Elaboración propia

DIMENSIÓN ASIGNATURA

Tabla 18:
Dimensión Asignatura

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Código de la asignatura	Código interno de la asignatura	34
Nombre	Nombre de la asignatura	Base de Datos
Ciclo	Ciclo de la asignatura	IV
Plan	Plan de estudios de la asignatura	16

Fuente: Elaboración propia

DIMENSIÓN CARRERA

Tabla 19:
Dimensión Carrera

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Código de la carrera	Código interno de la carrera	02
Nombre	Nombre de la carrera	Ingeniería de Sistemas

Fuente: Elaboración propia

DIMENSIÓN CICLO

Tabla 20:
Dimensión Ciclo

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Código del ciclo	Código interno del Ciclo	I

Fuente: Elaboración propia

DIMENSIÓN PERIODO

Tabla 21:
Dimensión Periodo

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Código del Periodo	Código interno del Periodo	1
Año	Año del Periodo	2017

Fuente: Elaboración propia

3.6.6.3 GRANULARIDAD

3.6.6.4 HECHOS

Tabla 22:
Tabla de Hechos

TABLA DE HECHOS	DESCRIPCIÓN
Información del rendimiento académico	Tabla de hechos referida al rendimiento académico de los alumnos

Fuente: Elaboración propia

3.6.6.5 MEDIDAS

Tabla 23:
Tabla de medidas

Medidas	Formulas
Asignaturas matriculadas por Curso	Número de Matriculados por curso
Asignaturas aprobadas	Número de cursos aprobados
Asignaturas desaprobadas	Número de cursos desaprobados
Créditos matriculados	Número de créditos matriculados
Créditos aprobados	Número de créditos aprobados
Créditos desaprobados	Número de créditos desaprobados
Promedio Ponderado	Promedio de curso por alumno

Fuente: Elaboración propia

3.6.7 DISEÑO FÍSICO

3.6.7.1 IDENTIFICACIÓN DEL MODELO TRANSACCIONAL

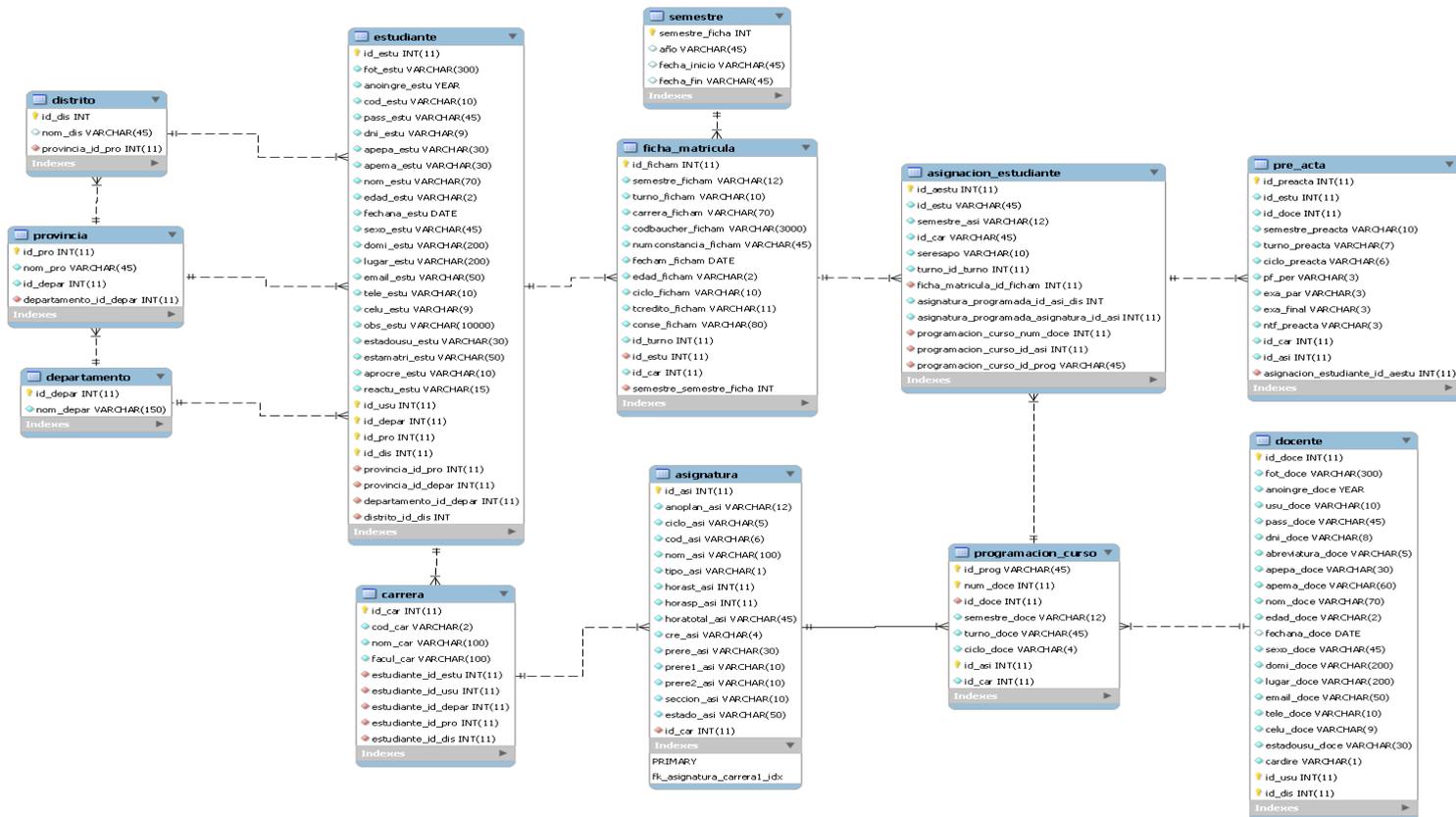


Figura 21: Modelo Transaccional

3.6.7.2 MODELO LÓGICO DEL DATAMART – MODELO ESTRELLA

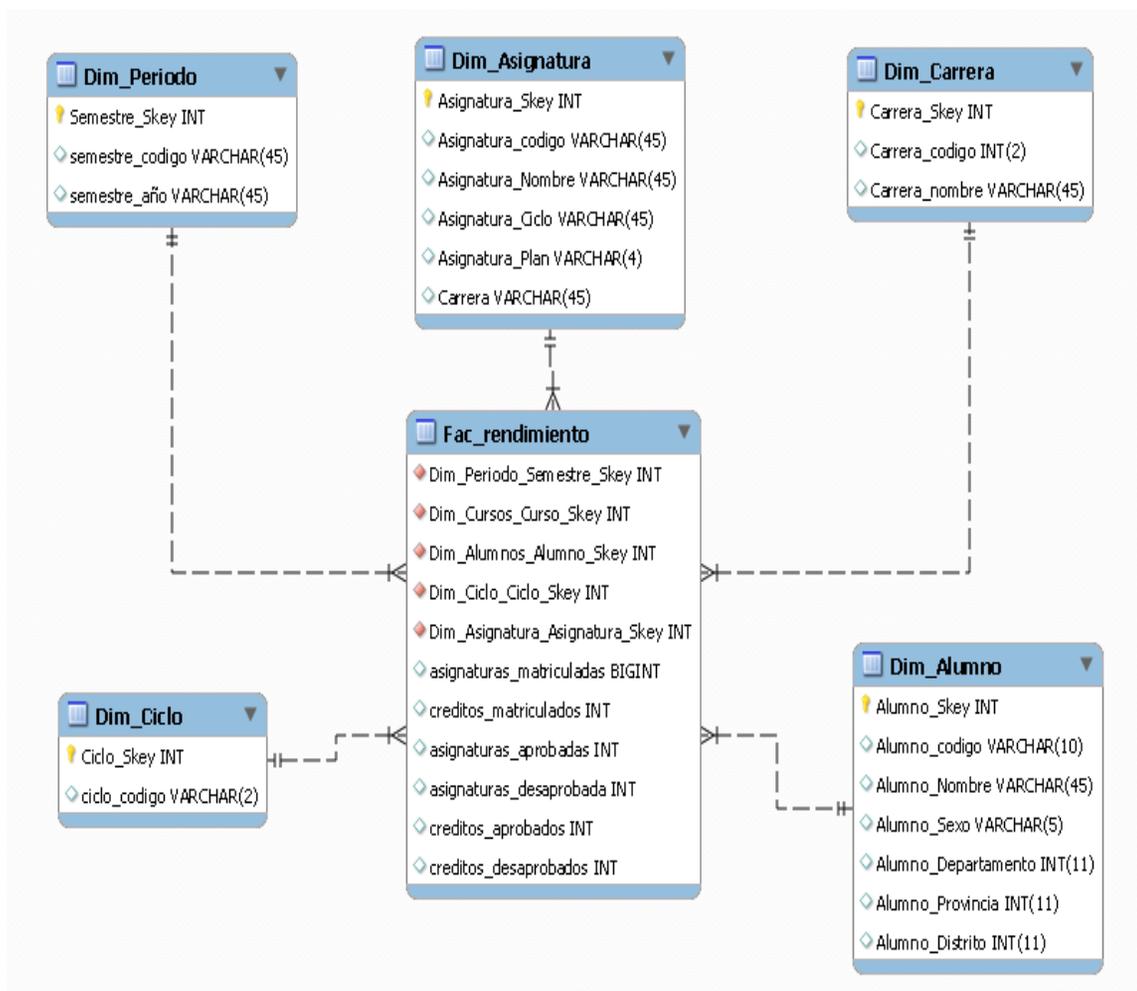


Figura 22: Modelo Estrella

3.6.7.3 DISEÑO FÍSICO

3.6.7.3.1 DICCIONARIO DEL DATAMART

DIMENSIÓN ALUMNO

Tabla 24:
Diccionario de datos de la dimensión Alumno

LLAVE	ATRIBUTO	TIPO	DESCRIPCIÓN
PK	Alumno_Skey	Int	Clave que identifica la alumno
	Alumno_codigo	Varchar(10)	Código del alumno
	Alumno Nombre	Varchar(100)	Nombre completo del contribuyente
	Alumno Sexo	Varchar(5)	Sexo del alumno
	Departamento	Int	Departamento de origen del alumno
	Provincia	Int	Provincia de origen del alumno
	Distrito	Int	Distrito de origen del alumno

Fuente: Elaboración propia

DIMENSIÓN ASIGNATURA

Tabla 25:
Diccionario de datos de la Dimensión Asignatura

LLAVE	ATRIBUTO	TIPO	DESCRIPCIÓN
PK	Asignatura_Skey	Int	Clave que identifica la alumno
	Asignatura_codigo	Varchar(10)	Código de la asignatura
	Asignatura_Nombre	Varchar(100)	Nombre completo del Alumno
	Asignatura_Ciclo	Varchar(5)	Sexo del alumno
	Asignatura_Plan	Int	Departamento de origen del alumno

Fuente: Elaboración propia

DIMENSIÓN CARRERA

Tabla 26:
Diccionario de datos de la Dimensión Carrera

LLAVE	ATRIBUTO	TIPO	DESCRIPCIÓN
PK	Carrera_Skey	Int	Clave que identifica la alumno
	Carrera_codigo	Varchar(10)	Código del alumno
	Carrera_Nombre	Varchar(100)	Nombre completo del contribuyente

Fuente: Elaboración propia

DIMENSIÓN CICLO

Tabla 27:
Diccionario de datos de la Dimensión Ciclo

LLAVE	ATRIBUTO	TIPO	DESCRIPCIÓN
PK	Ciclo_Skey	Int	Clave que identifica al ciclo
	Ciclo_codigo	Varchar(10)	Código del ciclo

Fuente: Elaboración propia

DIMENSIÓN PERIODO

Tabla 28:
Diccionario de datos de la Dimensión Periodo

ATRIBUTO	DESCRIPCIÓN	EJEMPLO
Código del Periodo	Código interno del Periodo	1
Año	Año del Periodo	2017

Fuente: Elaboración propia

3.6.8 DISEÑO E IMPLEMENTACIÓN DEL SUBSISTEMA DE ETL

3.6.8.1 CARGA DE DATOS A TABLAS DIMENSIONALES

Para el procedimiento de extracción y transformación de datos se crearon los siguientes flujos de datos.

Tabla 29:
Carga de Dimensione

flujos de datos:

DIMENSIÓN	FLUJO DE INFORMACIÓN
DIM_SEMESTRE	 DIM_SEMESTRE
DIM_ASIGNATURA	 DIM_ASIGNATURA
DIM_CARRERA	 DIM_CARRERA
DIM_CICLO	 DIM_CICLO
DIM_ESTUDIANTE	 DIM_ESTUDIANTE

Fuente: Elaboración propia

3.6.8.2 CARGA DE DATOS A TABLA DE HECHOS

Tabla 30:
Carga de Tabla de Hechos

TABLA DE HECHOS	FLUJO DE DATOS
FAC_RENDIMIENTOS_ACADEMICO	

Fuente: Elaboración propia

3.6.8.3 CARGA DE DIMENSIONES Y TABLA DE HECHOS

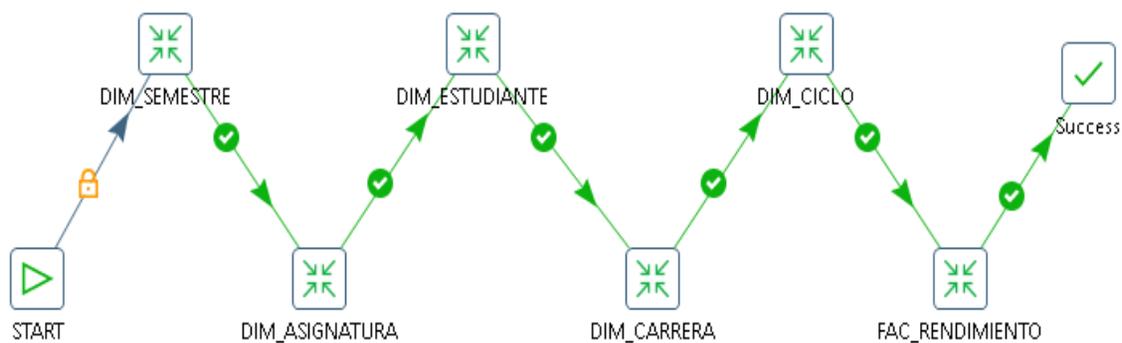


Figura 23: Modelo de carga de dimensiones y tabla de hechos

3.6.8.4 GESTIÓNAR CUBOS

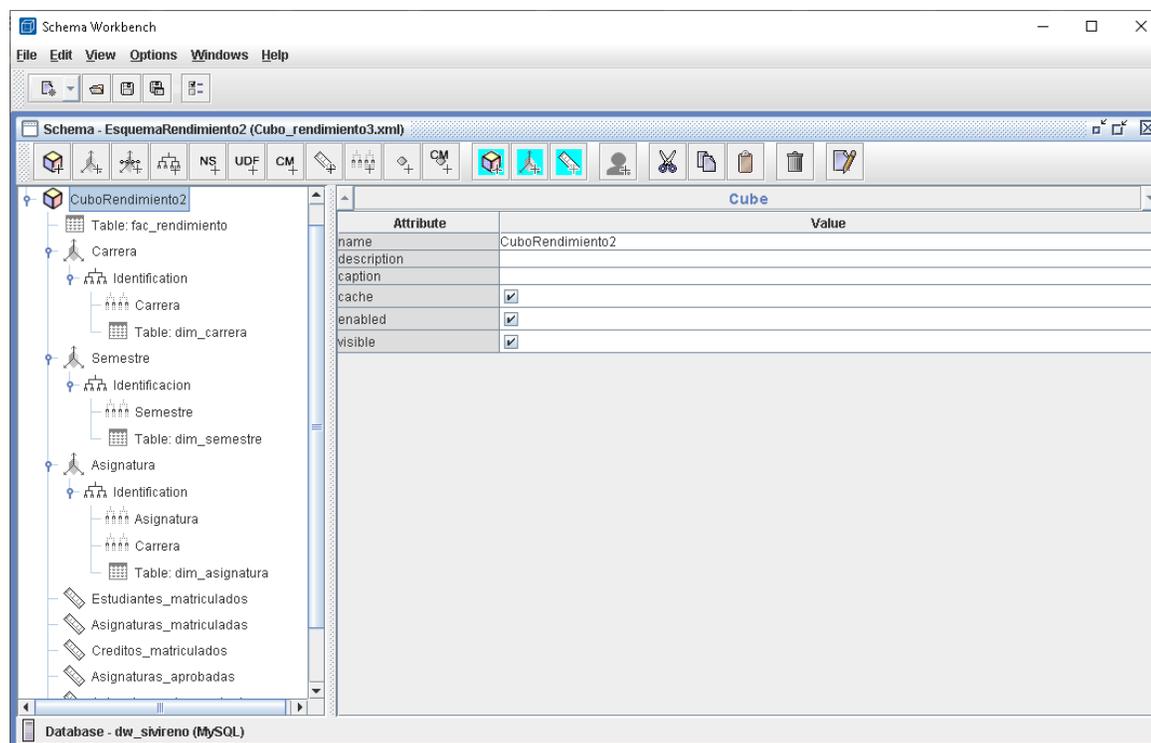


Figura 24: Modelo de Diseño de Cubo

3.6.9 ESPECIFICACIÓN DE APLICACIONES BI

Roles y procesos:

En esta fase se determinan los roles y permiso que se les asignará a los usuarios que tendrán acceso.

En este caso serán los siguientes:

Vicepresidente académico: El presente usuario tiene acceso reportes estratégicos para la toma de decisiones en el proceso académico

Directores de Escuela: Este usuario tiene acceso a la información relativa de sus Carreras Profesionales asignadas

3.6.10 DESARROLLO DE APLICACIONES DE BI

3.6.10.1 Reportes de JPivot

Con el fin de ir desarrollando los diferentes reportes utilizamos la herramienta Jpivot el cual nos permitirá la generación de Consultas MDX (MultiDimensional eXpression)

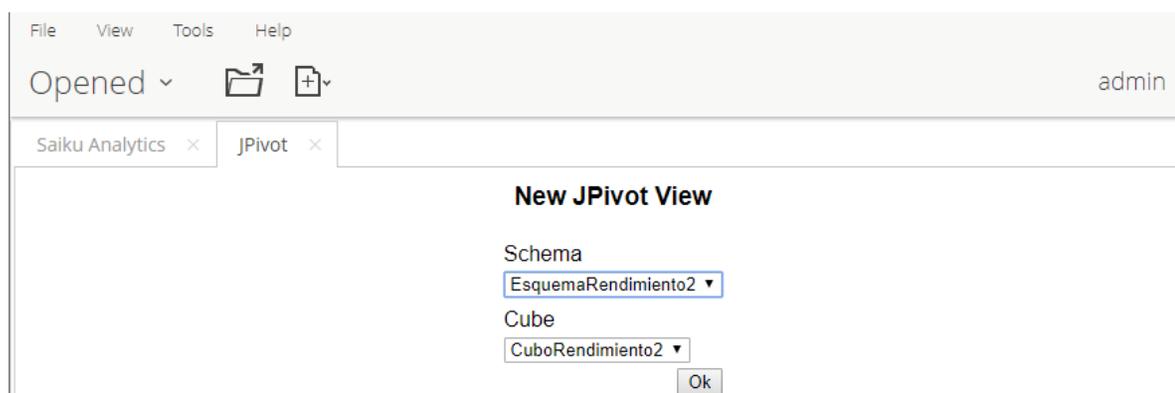


Figura 25: Interfaz del JPivot

CONSULTA MX ESTUDIANTES MATRICULADOS

```
select NON EMPTY {[Measures].[Estudiantes_matriculados]} ON COLUMNS, NON
EMPTY {([Carrera.Identification].[Todas las carreras], [Semestre.Identificacion].[Todos los
semestres], [Asignatura.Identification].[Todas las asignaturas])} ON ROWSfrom
[CuboRendimiento2]
```

CONSULTA MX ESTUDIANTES ASIGNATURAS_MATRICULADAS

```
select NON EMPTY {[Measures].[Estudiantes_matriculados]} ON COLUMNS, NON
EMPTY {([Carrera.Identification].[Todas las carreras], [Semestre.Identificacion].[Todos los
semestres], [Asignatura.Identification].[Todas las asignaturas])} ON ROWSfrom
[CuboRendimiento2]
```

CONSULTA MX ESTUDIANTES ASIGNATURAS_APROBADAS

```
select NON EMPTY {[Measures].[Asignaturas_aprobadas]} ON COLUMNS, NON
EMPTY {[Carrera.Identificacion].[Todas las carreras], [Semestre.Identificacion].[Todos los
semestres], [Asignatura.Identificacion].[Todas las asignaturas]} ON ROWSfrom
[CuboRendimiento2]
```

CONSULTA MX ESTUDIANTES ASIGNATURAS_DES APROBADAS

```
select NON EMPTY {[Measures].[Asignaturas_desaprobadas]} ON COLUMNS, NON
EMPTY {[Carrera.Identificacion].[Todas las carreras], [Semestre.Identificacion].[Todos los
semestres], [Asignatura.Identificacion].[Todas las asignaturas]} ON ROWSfrom
[CuboRendimiento2]
```

CONSULTA MX ESTUDIANTES CREDITOS APROBADOS

```
select NON EMPTY {[Measures].[Creditos_aprobados]} ON COLUMNS, NON EMPTY
{[Carrera.Identificacion].[Todas las carreras], [Semestre.Identificacion].[Todos los semestres],
[Asignatura.Identificacion].[Todas las asignaturas]} ON ROWSfrom [CuboRendimiento2]
```

CONSULTA MX ESTUDIANTES CREDITOS DESAPROBADOS

```
select NON EMPTY {[Measures].[Creditos_desaprobados]} ON COLUMNS, NON
EMPTY {[Carrera.Identificacion].[Todas las carreras], [Semestre.Identificacion].[Todos los
semestres], [Asignatura.Identificacion].[Todas las asignaturas]} ON ROWSfrom
[CuboRendimiento2]
```

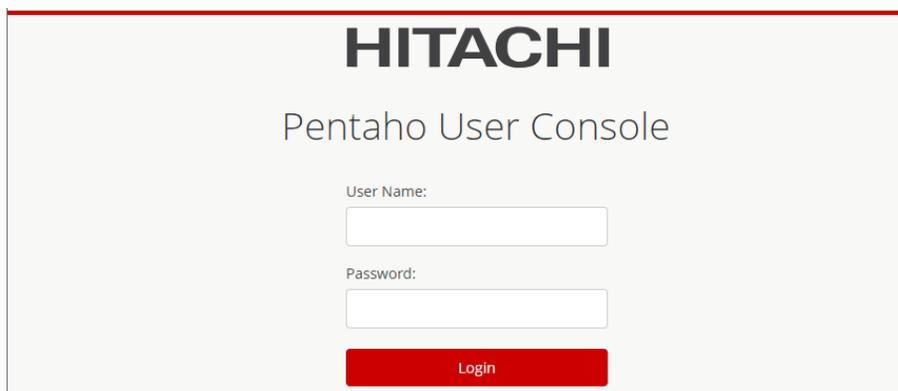
CONSULTA MX ESTUDIANTES PROMEDIO

```
select NON EMPTY {[Measures].[Creditos_promedio]} ON COLUMNS, NON EMPTY
{[Carrera.Identificacion].[Todas las carreras], [Semestre.Identificacion].[Todos los semestres],
[Asignatura.Identificacion].[Todas las asignaturas]} ON ROWSfrom [CuboRendimiento2]
```

3.6.11 IMPLEMENTACIÓN

3.6.11.1 ACCESO AL APLICATIVO WEB

Para ingresar al aplicativo se debe ingresar desde la intranet de la Universidad, donde el usuario debe ingresar sus datos de acceso al Sistema.



The image shows a login interface for the HITACHI Pentaho User Console. At the top, the word "HITACHI" is displayed in a large, bold, black font. Below it, the text "Pentaho User Console" is centered. There are two input fields: "User Name:" and "Password:". Below the password field is a red button with the text "Login" in white.

Figura 26: Pagina Inicial de Pentaho

3.6.11.2 REPORTE GENERADOS POR EL SISTEMA

A continuación, se muestra los distintos reportes desarrollados para la implementación de la metodología de Inteligencia de Negocios.

En el siguiente gráfico se muestra la cantidad de Asignaturas matriculadas por Semestre Académico.

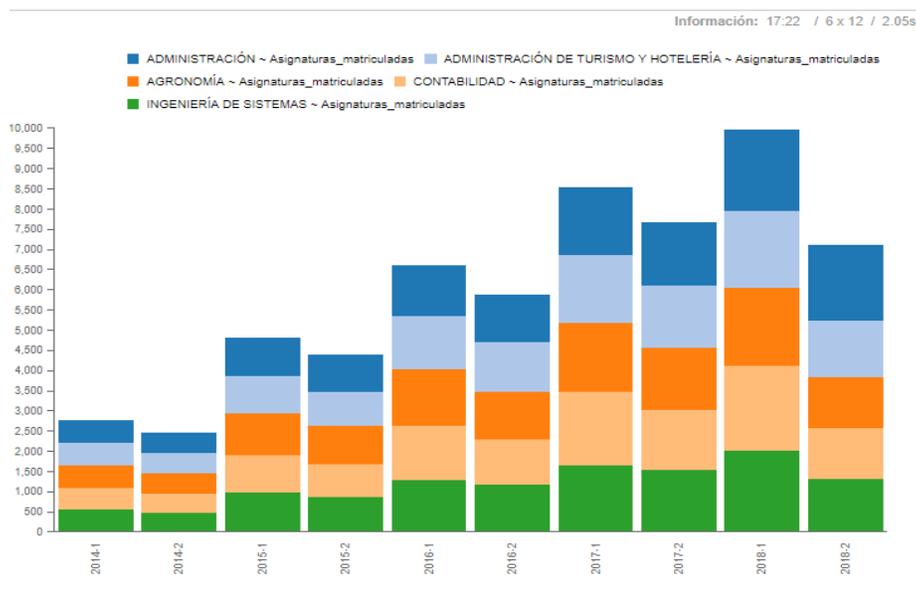


Figura 27: Asignaturas matriculadas

En el siguiente gráfico se muestra la cantidad de Asignaturas matriculadas por Semestre Académico en formato de Barras Verticales

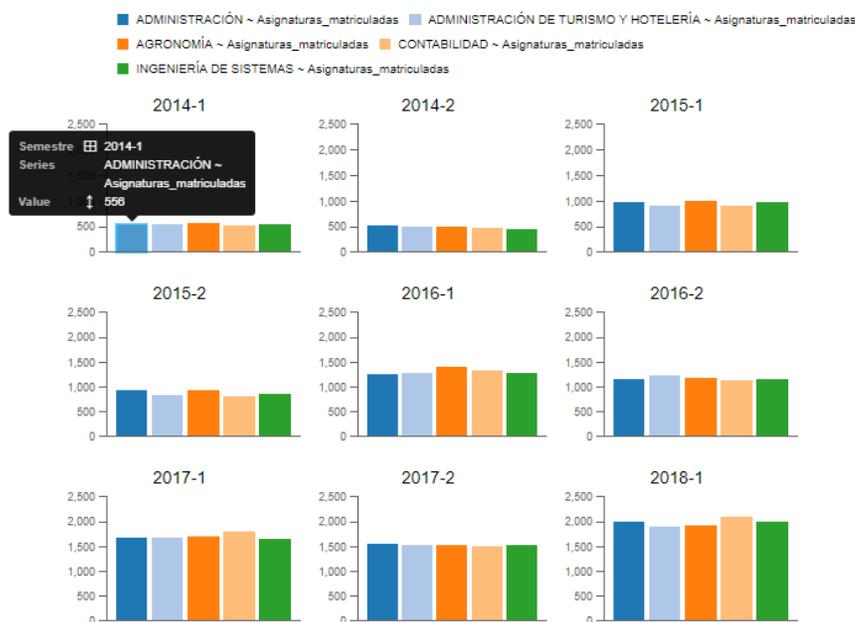


Figura 28: Asignaturas Matriculadas en Barra

En el siguiente gráfico se muestra la cantidad de Asignaturas aprobadas por Académico en Barras Múltiples

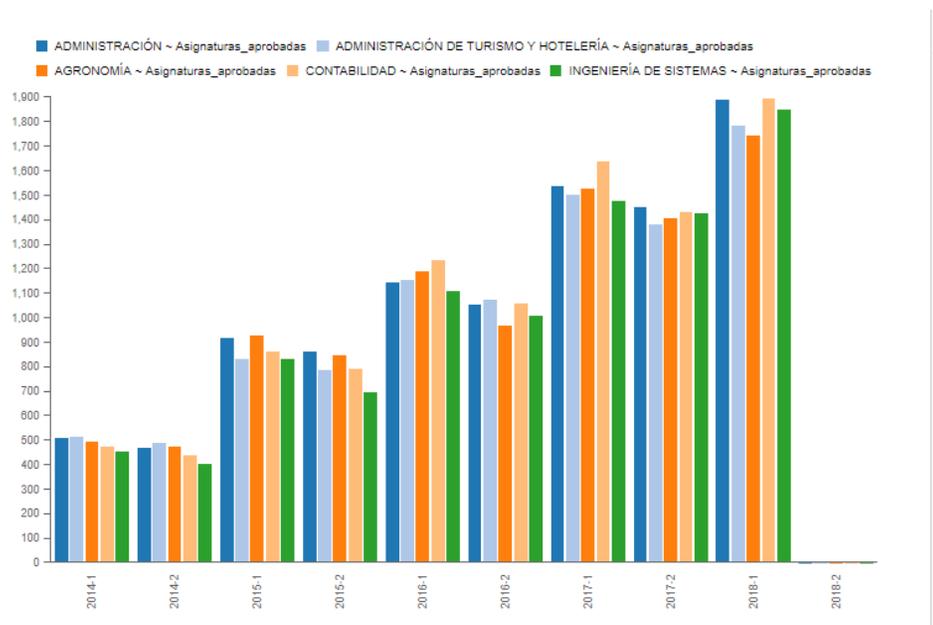


Figura 29: Asignaturas Aprobadas en Barras Múltiples

En el siguiente gráfico se muestra la cantidad de Asignaturas desaprobadas por Académico

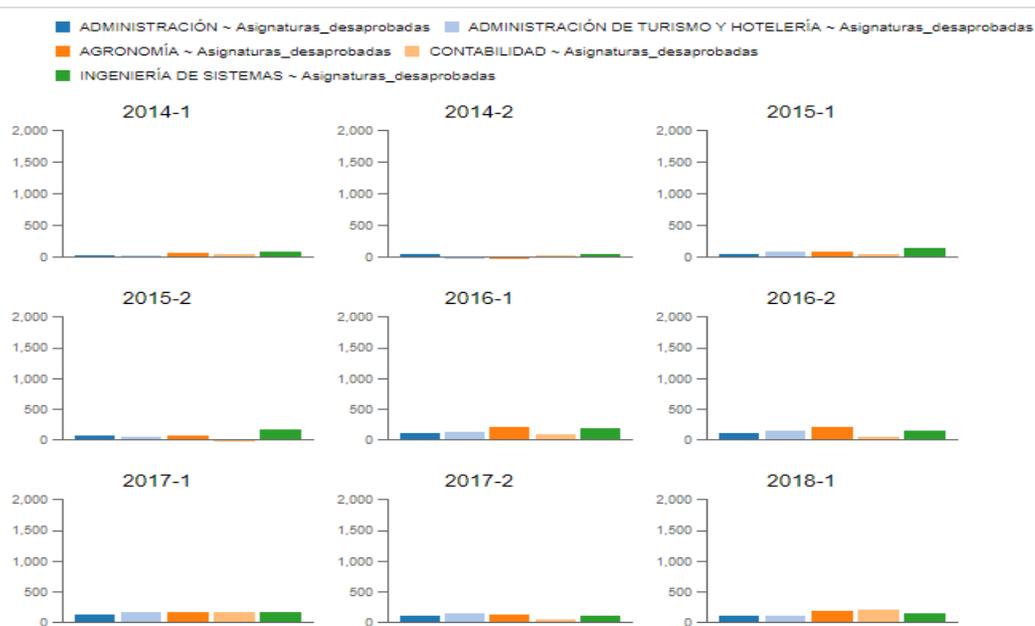


Figura 30: Asignaturas Desaprobadas por Semestre Académico

En el siguiente gráfico se muestra la cantidad de Créditos Matriculados por Semestre



Figura 31: Créditos Matriculados por semestre

En el siguiente gráfico se muestra la cantidad de Créditos Aprobados por Semestre



Figura 32: Créditos Aprobados por Semestre Académico

En el siguiente gráfico se muestra la cantidad de Créditos Desaprobados por Semestre

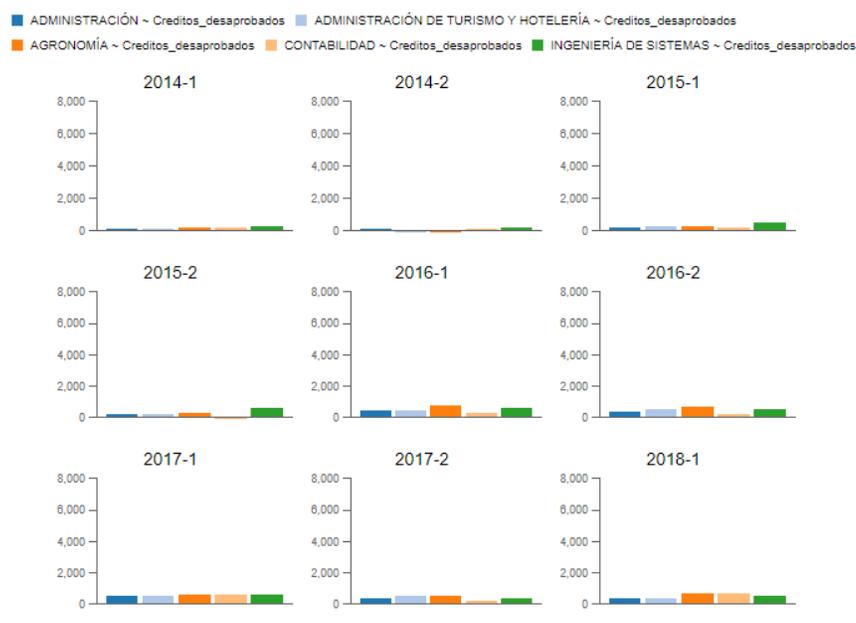


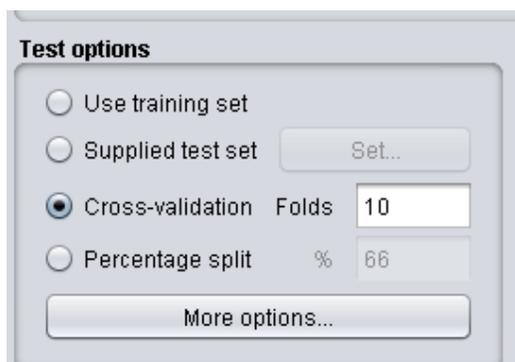
Figura 33: Créditos Desaprobados por Semestre

3.6.12 MINERÍA DE DATOS

3.6.12.1 RECOPIACION DE LA INFORMACION

Para la recopilación de información se ha obtenido del Datamart de rendimiento académico, por lo que hemos tomado una muestra de los alumnos ingresantes del semestre 2018-1 con el fin de probar nuestro modelo y para ello le aplicaremos el algoritmo de árboles de decisión J48

Configuramos le test utilizando el modo: Cross Validation



En la siguiente tabla se observa los resultados cuantitativos

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2110           89.4068 %
Incorrectly Classified Instances    250            10.5932 %
Kappa statistic                     0.0395
Mean absolute error                 0.1848
Root mean squared error             0.3055
Relative absolute error             96.0433 %
Root relative squared error         98.5902 %
Total Number of Instances          2360

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.999   0.976   0.895     0.999   0.944     0.121   0.601    0.924    APROBADO
                0.024   0.001   0.750     0.024   0.046     0.121   0.601    0.157    DESAPROBADO
Weighted Avg.   0.894   0.871   0.879     0.894   0.847     0.121   0.601    0.841

```

Figura 35: Resultados con Weka

A continuación, se observa los valores de la matriz de confusión

```

=== Confusion Matrix ===

      a    b  <-- classified as
2104    2 |   a = APROBADO
 248    6 |   b = DESAPROBADO

```

Figura 36: Matriz de Confusión

A continuación, se observa el árbol en formato de texto

```

MODALIDAD = ORDINARIO: APROBADO (1475.0/191.0)
MODALIDAD = CEPRE: APROBADO (568.0/14.0)
MODALIDAD = GRADUADOS Y TITULADOS
| SEXO = HOMBRE: DESAPROBADO (6.0)
| SEXO = MUJER: APROBADO (6.0)
MODALIDAD = VICTIMAS DE TERRORISMO: APROBADO (46.0/7.0)
MODALIDAD = PRIMEROS PUESTOS: APROBADO (213.0/17.0)
MODALIDAD = DISCAPACIDAD
| ESCUELA = AGRONOMÍA: APROBADO (0.0)
| ESCUELA = ADMINISTRACIÓN: APROBADO (12.0/4.0)
| ESCUELA = INGENIERÍA DE SISTEMAS: APROBADO (14.0/2.0)
| ESCUELA = ADMINISTRACIÓN DE TURISMO Y HOTELERÍA: DESAPROBADO (7.0)
| ESCUELA = CONTABILIDAD: DESAPROBADO (7.0/1.0)
MODALIDAD = TRASLADO EXTERNO: APROBADO (6.0)

Number of Leaves :    12

Size of the tree :    15

Time taken to build model: 0.01 seconds

```

Figura 37: Visualización del Árbol en modo Textual

A continuación, se observa el árbol de manera visual

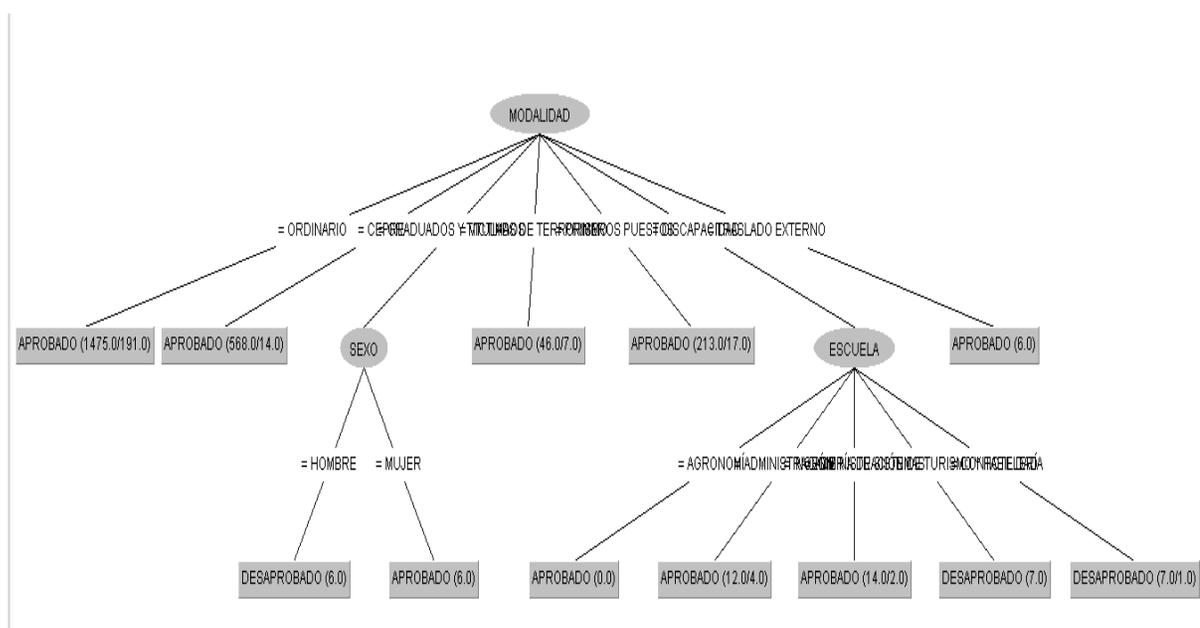


Figura 38: Visualización del Árbol

3.6.12.3 EVALUACION DE DESEMPEÑO

3.6.12.3.1 Análisis de la Matriz de Confusión

- Verdaderos positivos (TP) = Son el número de los estudiantes que SI pertenecen a la clase y el clasificador SI los definió en esa clase.
- Falsos positivos (FP) = Son el número de los estudiantes que NO pertenecen a la clase y el clasificador SI los definió en esa clase.
- Verdaderos negativos (TN)= Son el número de los estudiantes que NO pertenecen a la clase y el clasificador NO los definió en esa clase.
- Falsos negativos (FN) = Son el número de los estudiantes que SI pertenecen a la clase y el clasificador NO los definió en esa clase.

		Clase predicha	
		Aprobado	Desaprobado
Clase Verdadera	Aprobado	TP=2104	FP=2
	Desaprobado	FN=248	TN=6

Figura 39: Matriz de Confusión

$$Exactitud = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{2104+6}{2104+6+248+2} = 89.4068\%$$

Del valor anterior nos indica que tenemos un 89.4068% de instancias correctamente clasificadas.

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \times 100\% = \frac{248 + 2}{2104 + 6 + 2 + 248} = 10.5932\%$$

Del valor anterior nos indica que existe un error de clasificación del 10.5932%

3.6.12.3.2 SENSIBILIDAD

La sensibilidad nos indica la capacidad de nuestro clasificador para dar como casos aprobados los estudiantes realmente aprobados.

$$\text{Sensibilidad} = \frac{TP}{TP+FN} = \frac{2104}{2104+248} = 0.89$$

3.6.12.3.3 ESPECIFICIDAD

La especificidad es la capacidad del modelo predictivo de identificar una nueva instancia que pertenece a la clase desaprobado.

$$\text{Especificidad} = \frac{TN}{TN + FP} = \frac{6}{12} = 0.75$$

3.6.12.4 MODELO DE PREDICCIÓN

A continuación, se mostramos una muestra de las predicciones del modelo:

The screenshot shows a software interface with two main panels: 'Test options' and 'Classifier output'.

Test options:

- Use training set
- Supplied test set (Set...)
- Cross-validation Folds: 10
- Percentage split %: 66
- More options...

Classifier output:

inst#	actual	predicted	error	prediction
1	1: APROBADO	1: APROBADO	0.892	
2	1: APROBADO	1: APROBADO	0.892	
3	1: APROBADO	1: APROBADO	0.892	
4	1: APROBADO	1: APROBADO	0.892	
5	1: APROBADO	1: APROBADO	0.892	
6	1: APROBADO	1: APROBADO	0.892	
7	1: APROBADO	1: APROBADO	0.892	
8	1: APROBADO	1: APROBADO	0.892	
9	1: APROBADO	1: APROBADO	0.892	
10	1: APROBADO	1: APROBADO	0.892	
11	1: APROBADO	1: APROBADO	0.892	
12	1: APROBADO	1: APROBADO	0.892	
13	1: APROBADO	1: APROBADO	0.892	
14	1: APROBADO	1: APROBADO	0.892	
15	1: APROBADO	1: APROBADO	0.892	
16	1: APROBADO	1: APROBADO	0.892	
17	1: APROBADO	1: APROBADO	0.892	
18	1: APROBADO	1: APROBADO	0.892	
19	1: APROBADO	1: APROBADO	0.892	
20	1: APROBADO	1: APROBADO	0.892	
21	1: APROBADO	1: APROBADO	0.892	
22	1: APROBADO	1: APROBADO	0.892	
23	1: APROBADO	1: APROBADO	0.892	
24	1: APROBADO	1: APROBADO	0.892	
25	1: APROBADO	1: APROBADO	0.892	
26	1: APROBADO	1: APROBADO	0.892	
27	1: APROBADO	1: APROBADO	0.892	

Result list (right-click for options):

- 01:27:04 - trees.J48
- 09:25:00 - trees.J48

Figura 40: Muestra de predicciones del modelo

3.6.12.5 ANALISIS DEL MODELO

El valor de la precisión corresponde al 89.4068%. El tiempo para computar el modelo fue de 0.01 segundo. Las instancias correctamente clasificadas son 2110 que nos da una precisión del 89.4068%, mientras que los clasificados incorrectamente son 250 que nos da el 89.4068% restante. Se observa en la Figura 37 presenta un conjunto de reglas generadas, con un tamaño de 12 nodos y 15 hojas. Además de ellos se observa que las variables más importantes para predecir el rendimiento académico de un estudiante son en orden jerárquico: modalidad de ingreso, sexo y escuela profesional y la variable que no influyen son: distrito y asignatura

3.7 ANÁLISIS DE DATOS

3.7.1 RESULTADOS GENERICOS

- A) Planificación del Proyecto
 - Visión del Producto
 - Objetivo del Proyecto
 - Alcance del Proyecto
 - Equipos de Trabajo
 - Hardware y Software
 - Análisis de Riesgo
- B) Definición de Requerimientos del Negocio
 - Entrevista
 - Lista de requerimientos
- C) Diseño de la Arquitectura
 - Diseño de la Arquitectura Tecnológica
- D) Selección de productos e instalación
 - Rol de los productos de software
- E) Modelado Dimensional
 - Medidas
 - Dimensiones

- Granularidad
- F) Diseño Físico
- Identificación del modelo transaccional
 - Modelo lógico del Datamart
 - Diccionario del Datamart
- G) Diseño e Implementación del SubSistema de ETL
- Carga de datos a tablas dimensionales
 - Carga de datos a tablas de hechos
 - Carga de dimensiones y tablas de hechos
 - Diseño de Cubos
- H) Especificación de Aplicaciones de BI
- I) Desarrollo de Aplicaciones d BI
- J) Minería de Datos
- K) Implementación

3.7.2 RESULTADOS ESPECIFICOS

A continuación, se observan los valores de la Post-Prueba O₂ y Post-Prueba O₁:

TABLA N°32:

Registro de datos de Pre-Prueba O₁ y Post-Prueba O₂ para los KPI₁, KPI₂, KPI₃.

N	KPI-1: Tiempo de generación de reportes (segundos).		KPI-2: Tiempo de Análisis de la Información (minutos)		KPI-3: Grado de Satisfacción del uso del usuario	
	Pre-Prueba	Post-Prueba	Pre-Prueba	Post-Prueba	Pre-Prueba	Post-Prueba
1.	66	4	88	26	Malo	Excelente
2.	27	4	95	31	Malo	Bueno
3.	28	4	95	32	Malo	Excelente
4.	11	3	85	32	Regular	Excelente
5.	55	5	89	34	Malo	Excelente
6.	103	3	91	27	Malo	Excelente
7.	102	4	86	28	Regular	Bueno
8.	70	4	91	26	Malo	Excelente
9.	108	4	89	26	Malo	Excelente
10.	74	4	91	26	Malo	Excelente
11.	62	4	92	28	Regular	Excelente
12.	109	3	93	33	Malo	Bueno
13.	85	3	93	26	Bueno	Excelente
14.	107	5	90	34	Malo	Excelente
15.	65	3	88	32	Malo	Excelente
16.	118	4	88	28	Malo	Bueno
17.	119	3	87	30	Regular	Excelente
18.	84	4	92	30	Malo	Bueno
19.	97	5	86	26	Malo	Excelente
20.	62	4	91	29	Malo	Bueno
21.	88	4	90	26	Malo	Bueno
22.	119	4	86	35	Malo	Excelente
23.	102	4	92	33	Malo	Excelente
24.	115	3	95	30	Regular	Bueno
25.	108	4	92	27	Malo	Excelente
26.	96	5	86	25	Regular	Excelente
27.	64	4	86	29	Malo	Excelente
28.	62	4	86	34	Malo	Excelente
29.	96	3	90	35	Bueno	Excelente
30.	116	5	86	35	Malo	Bueno

Fuente: Elaboración propia

3.7.3 RESULTADOS NUMÉRICOS

A continuación, se visualizan las medidas de los KPI'

TABLA N°33:
Promedio de los Indicadores O₂ y O₁.

Indicador	Pre-Prueba (O ₁) (media:x1)	Post-Prueba (O ₂) (media:x2)	Comentario
KPI 1: Tiempo de generación de reportes.	83.93 s.	3.9 s.	-
KPI 2: Tiempo de análisis de la información.	89.633 m	29.77 m	-
KPI 3: Nivel de Satisfacción	-	-	No contrastado. Indicador Cualitativo

Fuente: Elaboración propia

4. CAPITULO IV: RESULTADOS

4.2 CÓNTRASTACIÓN DE HIPOTESIS

Para el procedimiento de la prueba de hipótesis se tomaron los siguientes parámetros:

- El nivel de confianza será de 95%
- El nivel de significancia será 5%.

a) Prueba para el Indicador. Tiempo de generación de reportes: KPI1

Hipótesis Específica H_{i1} :

Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del tiempo de generación de reportes del proceso de toma de decisiones del rendimiento académico (Post-Prueba O_2) con respecto a la muestra a que no se le aplico el estímulo (Pre-Prueba O_1).

Solución:

A. Planteamiento de la Hipótesis:

μ_1 = Media del tiempo de generación de reportes en la Pre-Prueba O_1 .

μ_2 = Media del tiempo de generación de reportes tes en la Post-Prueba O_2 .

H_0 : $\mu_1 = \mu_2$

H_a : $\mu_2 < \mu_1$

B. Criterio de Decisión:

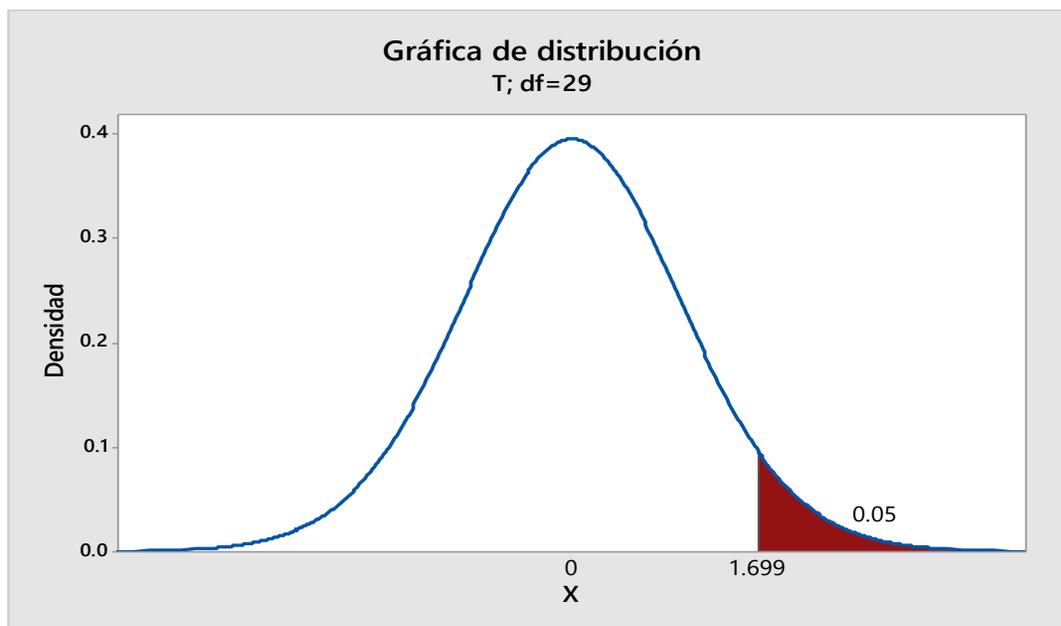


Figura 41: Criterio de decisión de KPI₁

C. Cálculo: Prueba t para medias de las dos muestras

TABLA N°34:
Resumen de Prueba t Student de KPI₁.

KPI ₁	Pre-Prueba O ₁	Post-Prueba O ₂
Media (x)	83.9 s.	3.9 s.
Desviación Estándar (S)	29.2	0.662
Observaciones (n)	30	30
Estimación de la diferencia	80.3	
Valor T: t _c	15	
Valor de p	0.000	
Valor crítico de t _{α/2} (una cola): t _t	1.699	

Fuente: Elaboración propia

D. Decisión Estadística:

Puesto que el valor- $p = 0.000 < \alpha = 0.05$, los resultados nos brindan suficiente información que rechaza la hipótesis nula (H_0), y aceptar la hipótesis alternativa (H_a).

La prueba es significativa con un nivel de confianza de 95%.

b) Prueba para el indicador: Tiempo de análisis de la información – KPI₂

Se valida la influencia de la implementación de la metodología de inteligencia de negocios, llevado a cabo en la muestra. Se realizó una medición antes de la implementación de la metodología de inteligencia de negocios (Pre-Prueba O_1) y otra después de la implementación de la metodología de inteligencia de negocios (Post-Prueba O_2).

Hipótesis Específica H_{12} : Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del tiempo de análisis de la información del proceso de toma de decisiones del rendimiento académico (Post-Prueba O_2) con respecto a la muestra a que no se le aplicó el estímulo (Pre-Prueba O_1).

Solución:

A. Planteamiento de la Hipótesis:

μ_1 = Media del tiempo de análisis de la información en la Pre-Prueba O_1 .

μ_2 = Media del tiempo de análisis de la información en la Post-Prueba O_2 .

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_2 < \mu_1$$

B. Criterio de Decisión:

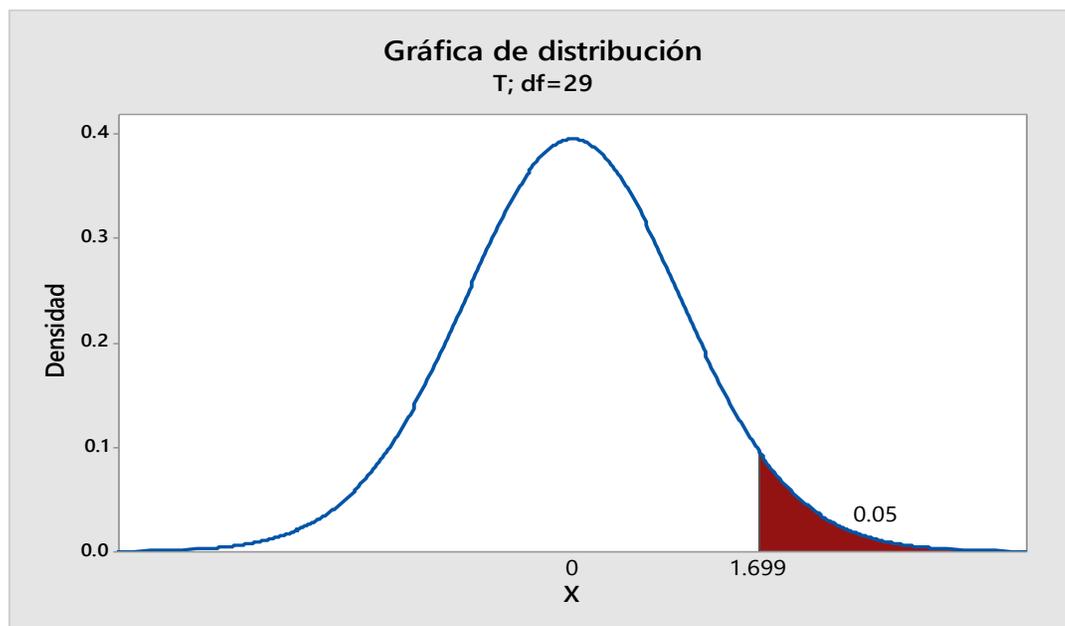


Figura 42: Distribución de Probabilidad del KPI 2.

C. Cálculo: Prueba t para medias de las dos muestras:

TABLA N°35:
Resumen de Prueba t Student de KPI₂.

KPI ₂	Pre-Prueba O ₁	Post-Prueba O ₂
Media (x)	89.63 m.	29.77 m.
Desviación Estándar (S)	3.045	3.320
Observaciones (n)	30	30
Diferencia hipotética de las medias	59.867	
Valor de T: t_c	72.68	
Valor p (una cola)	0.000	
Valor crítico de t_{α/2} (una cola): t_t	1.690	

Fuente: Elaboración propia

A. Decisión Estadística:

Puesto que el valor $p = 0.000 < \alpha = 0.05$, los resultados nos brindan suficiente información que rechaza la hipótesis nula (H_0), y aceptar la hipótesis alternativa (H_a).

La prueba es significativa con un nivel de confianza de 95%.

4.3 ANÁLISIS E INTERPRETACIÓN DE LA ESTADISTICA DESCRIPTIVA

En las siguientes Tablas se muestra los resultados de la Pre-Prueba O_1 y Post- Prueba O_2 .

a) Indicador. Tiempo de generación de reportes: KPI_1

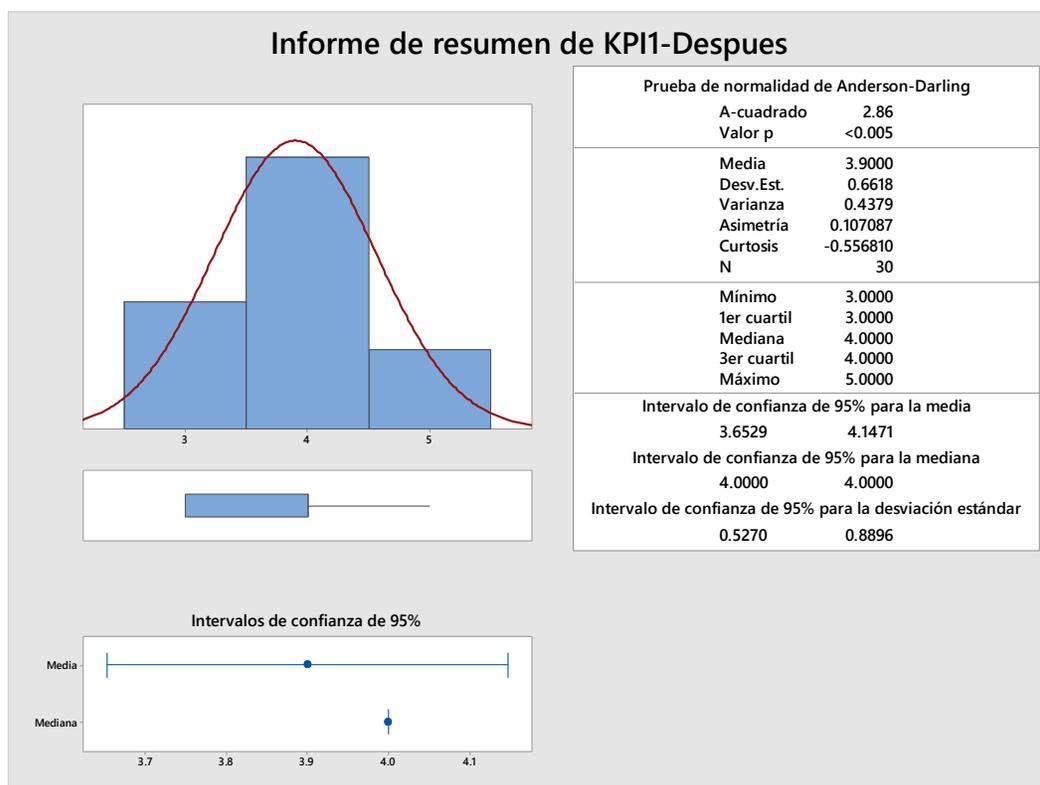


Figura 43: Resumen de Post-Prueba O_2 para KPI_1 .

- La media de tiempo de generación de es de 3.9 s.
- Alrededor del 95 % de tiempos de registro de la asistencia están entre 3.6569 s. y 4.1471 s.
- La curtosis= -0.556810 indica que es más plana la curva y es del tipo platocurtica
- La asimetría = 0.107087 indica que tiene asimetría positiva y que más valores agrupados hacia la izquierda de la curva, es decir que la mayoría de los tiempos de generación de reportes son bajos.
- El 1er Cuartil (Q1) = 3 indica que el 25% de los tiempos de generación de reportes es mayor que o igual a este valor.
- El 3er Cuartil (Q3) = 4 indica que el 75% de los tiempos de generación de reportes es mayor que o igual a este valor.

b) **Indicador. Tiempo de análisis de la Información: KPI₂**

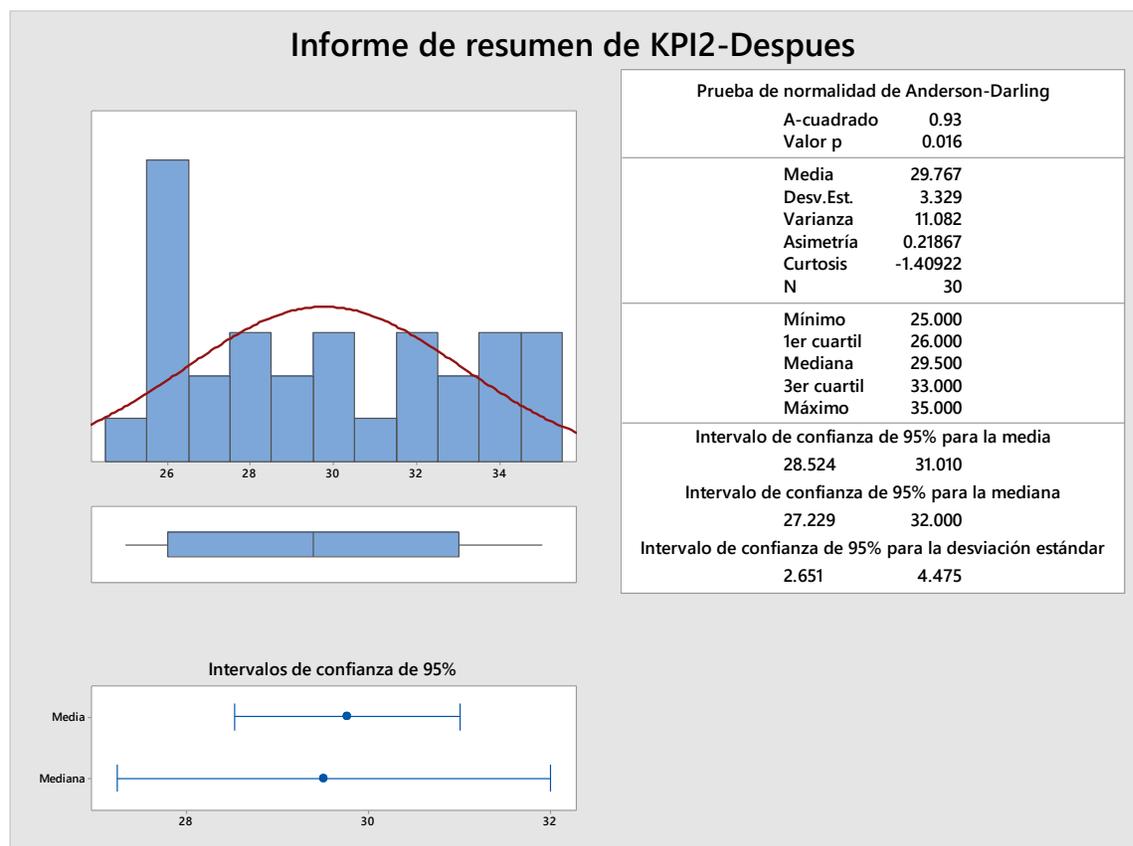


Figura 44: Resumen para KPI 2 de Post-Prueba O₂.

- La media del tiempo de análisis de la información es de 29.67 min.
- Alrededor del 95 % del tiempo de análisis de la información están entre 27.229 y 31.010
- La Kurtosis= -1.40922 indica que es más plana la curva y es del tipo platicurtica.
- La Asimetría = 0.21867 indica que la mayoría de los tiempos de análisis de la información son bajos.
- El 1er Cuartil (Q1) = 26 indica que el 25% de los es mayor que o igual a este valor.
- El 3er Cuartil (Q3) = 33 indica que el 75% de los tiempos de análisis de la información es menor que o igual a este valor.

c) **Indicador. Grado de satisfacción del uso de la aplicación Móvil: KPI₃**

TABLA N°36:
Valores de Pre-Prueba O₁ para KPI 3.

Estado	Frecuencia
Muy Mala	9
Mala	10
Regularmente Buena	6
Buena	3
Muy Buena	2

Fuente: Elaboración propia

**Resumen KPI3 (Pre-Prueba):
Nivel de Satisfacción**

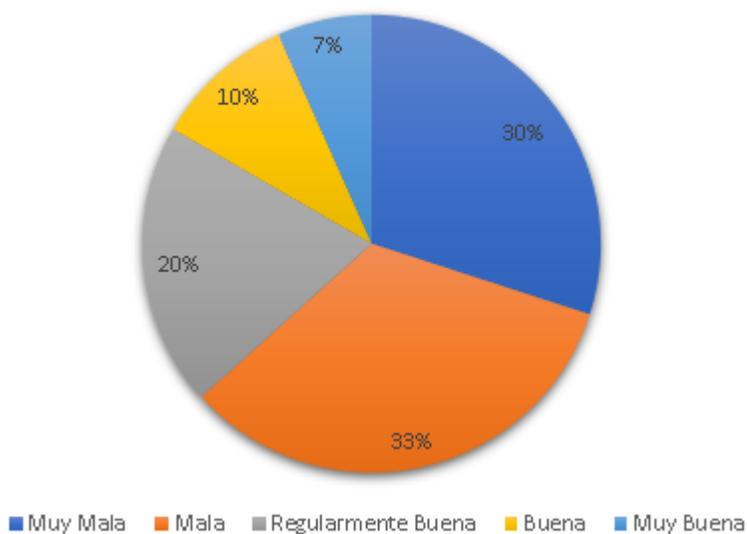


Figura 45: Resumen de Pre-Prueba O₁ para KPI 3.

- El 30% de las veces el Nivel de Satisfacción fue registrada como “muy mala”.
- El 33% de las veces el Nivel de Satisfacción fue registrada como “mala”.
- El 20% de las veces el Nivel de Satisfacción fue registrada como “regularmente buena”.
- El 10 % de las veces el Nivel de Satisfacción fue registrada como “buena”.
- El 7 % de las veces el Nivel de Satisfacción fue registrada como “muy buena”.

TABLA N°37:
Valores de Post-Prueba O₂ para KPI 3.

Estado	Frecuencia
Muy Mala	1
Mala	2
Regularmente Buena	2
Buena	8
Muy Buena	17

Fuente: Elaboración propia

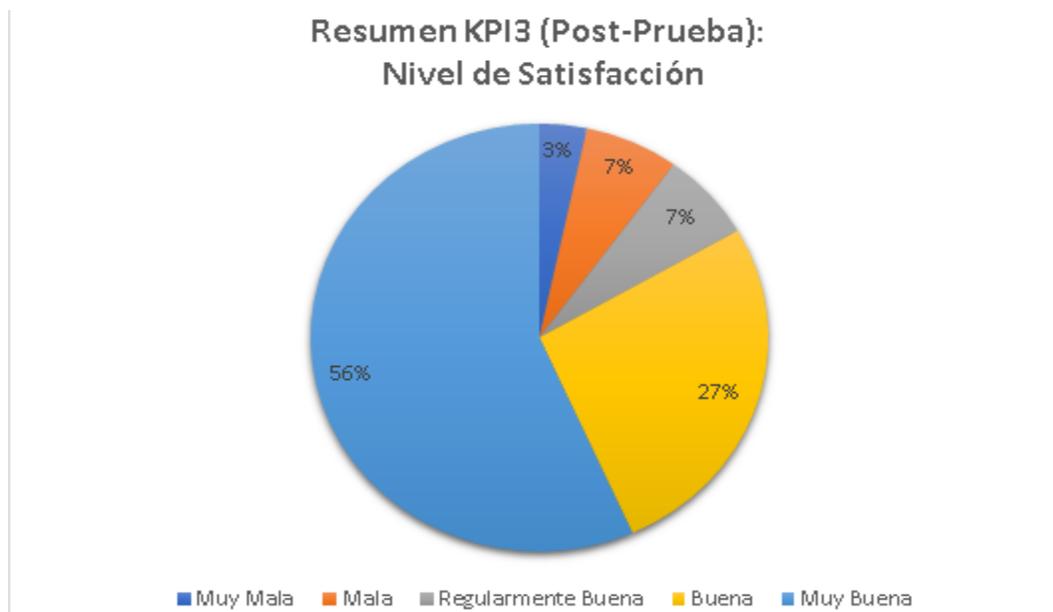


Figura 46: Resumen de Post-Prueba O₂ para KPI 3.

- El 3% de las veces el Nivel de Satisfacción fue registrada como “muy mala”.
- El 7% de las veces el Nivel de Satisfacción fue registrada como “mala”.
- El 7% de las veces el Nivel de Satisfacción fue registrada como “regularmente buena”.
- El 27 % de las veces el Nivel de Satisfacción fue registrada como “buena”.
- El 56 % de las veces el Nivel de Satisfacción fue registrada como “muy buena”.

5. DISCUSIÓN DE RESULTADOS

5.1 DISCUSION

El proceso de comprobación de la hipótesis nos ha permitido demostrar que efectivamente que la implementación influye positivamente en el proceso de toma de decisiones del proceso académico lo que coincide con la investigación DeSá Mussa, M., de Souza, S. C. & da Silva Freire, E. F (2018) llegaron a la conclusión que el uso del BI y la suite de Pentaho les ayudo en la toma de desiciones asi como tener una gran vistas y cubos disponible para los gerentes que les permite la explotación como factor desicivo para la toma de desiciones relevantes para el entorno educativo. Es importante la fase depuración de los datos el cual toma tiempo para llegar a consistenciar los datos, tal como Acosta, La Red Martínez & Primorac (2018), en su artículo denominado: “Determinación de perfiles de rendimiento académico en la UNNE con Minería de Datos Educacional” les tomo una gran demanda de tiempo de depuración y preparación debido a la inconsistencia inicial y finalmente se permitió crear un modelo predictivo para soportar la toma de decisiones eficaces. Zambrano (2017), en su Tesis: “Implementación de una solución de inteligencia de negocios acerca de la información de los docentes, estudiantes y personal administrativo de la Universidad Técnica del Norte para el Instituto de Altos Estudios” logro la implementación de la solución de inteligencia de negocios de acuerdo a los requerimientos del Instituto. En relación al objetivo específico relacionado al tiempo de generación de reportes en la presente investigación obtuvimos una mejora del 95%. La investigación de Peñaloza y Diego (2015), en sus tesis: “Implementación de un Sistema de Soporte de Decisiones para la Gestión Académica de la Universidad Nacional José María Arguedas” en su tesis demostraron la diferencia significa en los tiempos de demora, lo que significa que el presente trabajo de investigación también coincide en su demostración. En la presente tesis realizamos utilizamos la prueba de hipótesis de t-student tal como realizó, Duran (2017), en su tesis titulada “DataMart como parte de una solución de inteligencia de negocios, para el soporte de la toma de decisiones de la gestión académica de la Facultad de Ciencias de la USAM”, para la contratación de la hipótesis se utilizó la prueba de hipótesis de t-student donde cuya conclusión arribada fue, la implementación de un Data Mart, como parte de una solución de inteligencia de negocio, se obtuvo una mejora significativamente el proceso de Toma de Decisiones.

5.2 CONCLUSIONES

- a) Se logro influenciar reduciendo en un 95 % en el tiempo de generación de reporte relacionado con el proceso de toma de decisiones del rendimiento académico, lo cual se observa que la implementación de Inteligencia de negocios influyo positivamente en la mejora el tiempo de generación de reportes de 83.93 s. a 3.9 s. lo que permitió tomar decisiones en los tiempos adecuados.
- b) Se logro influenciar aumentando en un 67% el tiempo de análisis de información del proceso de toma de decisiones del rendimiento académico, lo cual se observa que la implementación de Inteligencia de negocios influyo positivamente en mejorar el nivel de precisión de la información de 89.633 m. a 29.77 m.
- c) Se logro influencias mejorando el nivel de satisfacción en un 87 % relacionado con proceso de toma de decisiones del rendimiento académico, lo cual se observa que la implementación de inteligencia de negocios influyo positivamente en mejorar el nivel de satisfacción de 7% a 56%, lo que permitió aumentar el nivel de satisfacción.

5.3 RECOMENDACIONES

- a) En relación a la primera conclusión se recomienda adquirir nueva infraestructura tecnológica que permita mejorar aún más el tiempo de procesamiento en la generación de los reportes
- b) En relación a la segunda conclusión se recomienda con el fin de disminuir el tiempo del proceso, realizar reuniones con los tomadores de decisiones para obtener nuevos requerimientos
- c) En relación a la tercera conclusión se recomienda con el fin de aumentar la satisfacción, poder realizar capacitaciones frecuentes de la herramienta de tal forma que puedan beneficiarse de sus bondades.

6. REFERENCIAS

- Acosta, J., La Red Martínez, D., & Primorac, C. (2018). Determinación de perfiles de rendimiento académico en la UNNE con minería de datos educacional. Argentina.
- Arenas López, M., & Gómez Montes, A. (2017). *Inteligencia de negocios aplicada a los procesos de autoevaluación de la Universidad de Manizales*.
- Ayala, D., & Omar, J. (2018). *Construcción de un Data Mart para la Gestión Académica de la Universidad Técnica del Norte (Bachelor's thesis)*. Ecuador.
- Duran Colonia, E. (2017). *Data Mart como parte de una solución de inteligencia de negocios, para el soporte de la toma de decisiones de la gestión académica de la Facultad de Ciencias en la UNASAM*.
- García Serrano, A. (2014). *Inteligencia Artificial*. México: Alfaomega.
- Giovaninni, M., Karanik, M., Pinto, N., & La Red Martínez, D. (2015). *Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos*. Argentina.
- Girones, J., Casas Roma, J., Minguillon Alfonso, J., & Caihuelas Quile, R. (2017). *Minería de Datos Modelos y algoritmos*. Barcelona: UOC.
- Holguín, C., & Monserrate, O. (2015). *Análisis, Diseño y Desarrollo del Datamart actividades de docentes y estudiantes para el Prototipo de Sistema Académico Integrado en las Carreras de Ingeniería en Sistemas Computacionales e Ingeniería en Networking y Telecomunicaciones de la Facultad de C*.
- Lozano Mejía, M. (2011). *Inteligencia de negocios: una opción ante la toma racional de decisiones responsables*.
- Pascal, G., Servetto, D., Unelén, L., & Yésica, L. (2017). Aplicación de Business Intelligence para la toma de decisiones en Instituciones Universitarias. Implementación de Boletines Estadísticos en la Universidad Nacional de Lomas de Zamora. *Revista Electrónica sobre Tecnología, Educación y Sociedad*.
- Peñaloza, R., & Diego, J. (2015). *Implementación de un Sistema de Soporte de Decisiones para la Gestión Académica de la Universidad Nacional José María Arguedas*.
- Rafael, D., & Elizabeth, L. (2017). *Implementación de un data mart para el seguimiento académico de los estudiantes en la escuela académico profesional de ingeniería de sistemas de la Universidad Nacional de Cajamarca*.

- Santoyo, J. (2018). *Modelo de datamart para análisis de indicadores de desempeño estudiantil en la SEPI UPIICSA*. México.
- Sarmiento, P. (2018). *Inteligencia de negocios usando Pentaho para la gestión académica en la UNAMBA-2016X*.
- Vázquez, J., Frías, R., González, L., & Farías, J. (2018). *ANÁLISIS DE INFORMACIÓN DE ASPIRANTES A INGRESAR A INSTITUCIONES DE EDUCACIÓN SUPERIOR UTILIZANDO PENTAHO*. México.
- Yamao, E. (2018). *Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de las Escuela Profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres*. Lima-Perú.
- Zambrano Segovia, E. (2017). *Implementación de una solución de inteligencia de negocios para la gestión del talento humano en la Universidad Técnica del Norte (Master's thesis)*.

7. ANEXOS

ANEXO 01: MATRIZ DE CONSISTENCIA

“METODOLOGÍA DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE TOMA DE DECISIONES DEL RENDIMIENTO ACADEMICO EN LA UNIVERSIDAD NACIONAL DE CAÑETE”

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	METODOLOGIA
<p>Problema general ¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete?</p>	<p>Objetivos General Determinar el grado de influencia de la metodología de inteligencia de negocios en el proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete.</p>	<p>Hipótesis General: Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del proceso de toma de decisiones del rendimiento académico en la Universidad Nacional de Cañete.</p>	<p>Variable independiente: Metodología de inteligencia de negocios</p>	<p>Tipo de Investigación: Investigación Aplicada</p>
<p>Problema Específico ¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el tiempo de generación de reportes del proceso de toma de decisiones del rendimiento académico?</p>	<p>Objetivos específicos: Determinar el grado de influencia de la de metodología de inteligencia de negocios en el tiempo de generación de reporte del proceso de toma de decisiones del rendimiento académico.</p>	<p>Hipótesis Específicas: H1: Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en la mejora del tiempo de generación de reportes del proceso de toma de decisiones del rendimiento académico.</p>	<p>Indicadores: Presencia-Ausencia</p>	<p>Diseño de Investigación: Pre-experimental</p>
<p>¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el tiempo de análisis de información del proceso de toma de decisiones del rendimiento académico?</p>	<p>Determinar el grado de influencia de la metodología de metodología de inteligencia de negocios en el tiempo de análisis de información del proceso de toma de decisiones del rendimiento académico.</p>	<p>H2: Si se aplica la metodología de inteligencia de negocios entonces influye positivamente en el tiempo de análisis de información del proceso de toma de decisiones del rendimiento académico.</p>	<p>Variable dependiente: Proceso de toma de decisiones del rendimiento académico de la Universidad Nacional de Cañete</p>	<p>Población: En el presente trabajo de investigación la población de estudio los procesos de Toma de decisiones de la gestión académica de la Universidad Nacional de Cañete</p>
<p>¿Cuál es el grado de influencia de la metodología de inteligencia de negocios en el nivel de satisfacción del usuario en el proceso de toma de decisiones del rendimiento académico?</p>	<p>Determinar el grado de influencia de la metodología de inteligencia de negocios en el nivel de satisfacción del usuario del proceso de toma de decisiones del rendimiento académico.</p>	<p>H3: Si se aplica la metodología de metodología de inteligencia de negocios entonces influye positivamente en la mejora del nivel de satisfacción del usuario del proceso de toma de decisiones del rendimiento académico.</p>	<p>Indicadores: Tiempo de generación de reportes. Tiempo de análisis de la información Nivel Satisfacción</p>	<p>Muestra: n=30 Toma de decisiones (*) *. Pande, P. Las claves Prácticas de SIX SIGMA, 2004, pp. 135-136.</p>

TABLA N°38: Matriz de Consistencia

ANEXO 02

PRESUPUESTO PROPUESTO

✓ Recursos Humanos

TABLA N°39:

Bienes utilizados en el desarrollo de la Tesis.

Descripción	Total
Desarrollo de la aplicación	S/.6000.00
TOTAL	S/.6000.00

✓ Bienes

TABLA N°40:

Bienes utilizados en el desarrollo de la Tesis.

Descripción	Costo	Cantidad/	Total
	unitario	horas	
Escritorio	280.00	1	280.00
Silla	80.00	1	80.00
Útiles de escritorio	20.00	10	200.00
USB 16G	40.00	1	40.00
Laptop	3000.00	1	3000.00
Impresora	250.00	1	250.00
TOTAL			S/.3850.00

Fuente: Elaboración Propia.

✓ **Servicios**

TABLA N°41:
Servicios utilizados en el desarrollo de la Tesis

Descripción	Costo unitario	Cantidad /	
		horas	Total
Internet	1.00	200 (horas)	200.00
Luz	150.00	280 (horas)	150.00
Pasajes	20.00	20.00	400.00
TOTAL			S/.750.00

Fuente: Elaboración Propia.

El costo total para el desarrollo e implementación de la solución es de S/.10000.00.

ANEXO 03
CUESTIÓARIO DE SATISFACCIÓN

Fecha: __/__/____ **Nombres:** _____

Instrucciones: Marque con un aspa la respuesta que crea conveniente teniendo en consideración el puntaje que corresponda de acuerdo al siguiente ejemplo: Muy Mala (1), Mala (2), Regularmente Buena (3), Buena (4), Muy Buena (5).

N°	Pregunta	Valoración				
		1	2	3	4	5
Metodología de inteligencia de negocios						
	Usabilidad					
1	Le parece fácil el uso del Sistema					
2	Piensa que el diseño del Sistema es fácil de usar					
	Disponibilidad					
3	El acceso a la información está disponible de manera ininterrumpida					
Proceso de toma de decisiones académicas						
	Satisfacción del usuario					
5	La implementación de metodología de inteligencia de negocios satisface sus necesidades de información					

ANEXO 04: Formato de Registro de KPI's

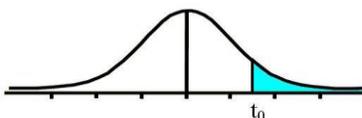
Tesista	Ricardo Carlos Inquilla Quispe
Organización	Universidad Nacional de Cañete
Ubicación	Jr. San Agustín No 124, San Vicente de Cañete Lima, Perú

Ítem	KPI-1		KPI-2	
	Pre-Prueba	Post-Prueba	Pre-Prueba	Post-Prueba
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				
18.				
19.				
20.				
21.				
22.				
23.				
24.				
25.				
26.				
27.				
28.				
29.				
30.				

ANEXO 05

TABLA T-STUNDET

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3007	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800

50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6458
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

ANEXO 06

GLOSARIO DE TERMINOS

- **Aplicación:** Programa que provee Funciones requeridas por un Servicio TI. Cada aplicación podría ser parte de más de un Servicio TI. Una Aplicación se puede ejecutar en uno o más Servidores o Clientes.
- **Dashboard:** Es una interfaz, tablero de mando, donde el usuario puede consultar la información de manera fácil.
- **Datawarehouse:** es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta.
- **Datamining:** El Data Mining es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos que expliquen el comportamiento de estos datos.
- **Dimensión:** Abreviatura de tecnología de la información.
- **ETL:** Es un lenguaje de programación de uso general de código del lado del servidor originalmente diseñado para el desarrollo web de contenido dinámico.
- **Indicadores:** Interfaz de Programación de Aplicaciones (Application Programming Interface) es el conjunto de funciones y procedimientos que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.
- **Inteligencia de negocios:** Integrated Development Environment, es una aplicación informática que proporciona servicios integrales para facilitarle al desarrollador o programador el desarrollo de software.
- **Indicador:** Es un dato o conjunto de datos que ayudan a medir objetivamente la evolución de un proceso o de una actividad.
- **Requerimiento:** Describen los servicios que ha de ofrecer el sistema y las restricciones asociadas a su funcionamiento.

- **KPI:** En inglés Key Performance Indicators, o Indicadores Clave de Desempeño, Miden el nivel del desempeño de proceso, centrándose en el "como" e indicando el rendimiento de los procesos, de forma que se pueda alcanzar el objetivo fijado.
- **Kurtosis:** Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, podemos identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) o una baja concentración (Platicúrtica).
- **Nivel de significación (σ):** Se define como la probabilidad de rechazar erróneamente la hipótesis nula.
- **Hipótesis alternativa:** Conocida como, es cualquier hipótesis que difiere de la hipótesis nula.
- **Hipótesis nula:** Se denomina hipótesis nula a la hipótesis que se desea contrastar.
- **Variable independiente:** Es la variable que el investigador mide, manipula o selecciona para determinar su relación con el fenómeno o fenómenos observados.
- **Variable Dependiente:** Son aquellas que teóricamente afectan a la variable dependiente pero no pueden medirse o manipularse. Normalmente son variables que se deducen de los efectos de las variables: independiente y moderador, sobre la variable dependiente