

“UNIVERSIDAD TÉCNICA DEL NORTE”



Facultad de Ingeniería en Ciencias Aplicadas
Carrera de Software

**ESTUDIO DE TÉCNICAS AVANZADAS DE ANÁLISIS DE DATOS PARA EL
PROCESO DE AUDITORÍA INFORMÁTICA EN EL MÓDULO DEL SISTEMA
ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE (UTN).**

Trabajo de grado previo a la obtención del título de Ingeniero en Software

Autor:

Francisco Javier Alvarez Osorio

Director:

MSc. Daisy Elizabeth Imbaquingo Esparza

Ibarra – Ecuador

2023



UNIVERSIDAD TÉCNICA DEL NORTE

BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN

A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	1050140266		
APELLIDOS Y NOMBRES:	ALVAREZ OSORIO FRANCISCO JAVIER		
DIRECCIÓN:	URCUQUÍ, CIUDADELA MUNICIPAL		
EMAIL:	fjalvarezo@utn.edu.ec		
TELÉFONO FIJO:	N/A	TELÉFONO MÓVIL:	0988098669

DATOS DE LA OBRA	
TÍTULO:	ESTUDIO DE TÉCNICAS AVANZADAS DE ANÁLISIS DE DATOS PARA EL PROCESO DE AUDITORÍA INFORMÁTICA EN EL MÓDULO DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE (UTN).
AUTOR (ES):	FRANCISCO JAVIER ALVAREZ OSORIO

FECHA: DD/MM/AAAA	14/09/2023
SOLO PARA TRABAJOS DE GRADO	
PROGRAMA:	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
TITULO POR EL QUE OPTA:	INGENIERO EN SOFTWARE
ASESOR /DIRECTOR:	MSc. Daisy Elizabeth Imbaquingo Esparza

2. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de esta y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, 14 de septiembre de 2023

EL AUTOR:



.....
Francisco Javier Alvarez Osorio

CERTIFICACION DIRECTOR

Ibarra, 14 de septiembre del 2023

CERTIFICACIÓN DEL DIRECTOR DEL TRABAJO DE GRADO

Por medio del presente yo MSc. Imbaquingo Esparza Daisy Elizabeth, certifico que el Sr. Francisco Javier Alvarez Osorio portador de la cedula de ciudadanía número 1050140266, ha trabajado en el desarrollo del proyecto de grado **“ESTUDIO DE TÉCNICAS AVANZADAS DE ANÁLISIS DE DATOS PARA EL PROCESO DE AUDITORÍA INFORMÁTICA EN EL MÓDULO DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE (UTN)”**, previo a la obtención del Título de Ingeniería en Software, esté trabajo se ha realizado con interés profesional y responsabilidad que certifico con honor de verdad.

Atentamente



MSc. Daisy Imbaquingo
DIRECTOR DE TRABAJO DE GRADO

Dedicatoria

Dedico este trabajo a mi madre Carmen y a mi padre Francisco, quienes me han aconsejado para alcanzar mis metas y supieron brindarme su apoyo desde el comienzo hasta el final de esta carrera.

A mis hermanos Samantha, Paul, Andrés y mi abuelita Aura por ser el impulso que me hace seguir avanzando para cumplir mis sueños.

Agradecimientos

Agradezco a mis padres por el esfuerzo y sacrificio que han hecho por mí, por ser la persona que soy gracias a su ejemplo y educación.

A toda mi familia, al estar siempre pendiente en cada paso que doy, por ser mi motivación y guiarme por el camino correcto.

A mis amigos y compañeros de carrera Leo, Leslie, Oliver, Marcelo, Jerson, Stiphen, Yamilex y Stalin por estar en cada instante ayudándome tanto en lo académico como personal.

A la MSc. Daisy Imbaquingo quien admiro mucho por tan trabajadora, por brindarme su tiempo y confianza en la realización de este trabajo y en cada una de las materias que fui su alumno.

Tabla de Contenido

Tabla de Contenido.....	7
Índice de Tablas.....	11
Índice de Figuras.....	13
Resumen.....	15
Abstract.....	16
Introducción.....	17
Antecedentes.....	17
Situación actual.....	17
Planteamiento del Problema.....	18
Objetivo General.....	19
Objetivos Específicos.....	19
Alcance y Metodología.....	19
Alcance.....	19
Metodología.....	20
Justificación y Riesgos.....	21
Justificación.....	21
CAPÍTULO I.....	24
1.1. Introducción.....	24
1.2. Señalamiento de Variables.....	24
1.3. Auditoría.....	25
1.4. Tipos de Auditoría.....	25
1.5. Auditoría Informática.....	26

1.5.1.	Áreas de la Auditoría Informática	27
1.6.	Proceso de Auditoría Informática	28
1.6.1.	Fase I: Planificación	30
1.6.2.	Fase II: Ejecución	31
1.6.3.	Fase III: Comunicación de Resultados	31
1.6.4.	Fase IV: Validación	32
1.6.5.	Fase V: Seguimiento	32
1.7.	Auditoría de Datos	32
1.7.1.	Etapas de Auditorías de Datos	33
1.7.2.	Calidad de Datos	34
1.8.	Procesamiento de datos	37
1.8.1.	Tipos de procesamiento de datos	37
1.8.2.	Técnicas de procesamiento	38
1.9.	Revisión de la literatura	39
1.9.1.	Preguntas de investigación	39
1.9.2.	Búsqueda de Documentos	40
1.10.	Técnicas de Procesamiento y Análisis de Datos	48
1.10.1.	Estadística Tradicional	49
1.10.2.	Ciencia de datos	50
1.11.	Técnicas Avanzadas de Procesamiento y Análisis de Datos	51
1.12.	Técnicas Avanzadas de Procesamiento y Análisis de Datos en Auditoría ..	54
1.12.1.	Selección de técnicas avanzadas	56
1.12.2.	Cuadrante de Gartner	59

1.13.	Metodología CRISP-DM	59
2.	CAPÍTULO II	63
2.1.	Aplicación del Metodología CRISP- DM	63
2.1.1.	<i>Comprensión del negocio</i>	63
2.1.2.	<i>Comprensión de los Datos</i>	65
2.1.3.	<i>Preparación de los datos</i>	66
2.1.4.	<i>Modelamiento</i>	67
2.1.5.	<i>Evaluación</i>	80
2.1.6.	<i>Despliegue</i>	80
2.2.	Planificación de Auditoría (Fase I).....	81
2.2.1.	<i>Contacto con el cliente.</i>	81
2.2.2.	<i>Estudio Inicial del Entorno a Auditar</i>	81
2.2.3.	<i>Equipo Auditor</i>	81
2.2.4.	<i>Plan Preliminar</i>	81
2.2.5.	<i>Propuesta de Auditoría</i>	85
2.2.6.	<i>Contrato.</i>	86
2.3.	Ejecución de la Auditoría (Fase II)	86
2.3.1.	<i>Plan de Trabajo</i>	87
2.3.2.	<i>Investigación de Campo</i>	90
2.3.3.	<i>Análisis de la información recolectada</i>	94
2.3.4.	<i>Conclusiones preliminares</i>	115
2.4.	Comunicación De Resultados (FASE III).....	116
2.4.1.	<i>Elaboración Del Informe Final Preliminar</i>	116

2.4.2. Informe Final Preliminar	116
3. CAPÍTULO III	122
3.1. Validación (FASE IV)	122
3.1.1. <i>Calidad</i>	122
3.1.2. <i>Seguridad de la información</i>	126
3.1.3. <i>Cumplimiento</i>	128
3.2. Seguimiento (FASE V).....	133
3.2.1. <i>Informe Final de Auditoría</i>	133
3.2.2. <i>Contacto con el Cliente</i>	133
Conclusiones	134
Recomendaciones	135
Referencias.....	136

Índice de Tablas

Tabla 1 Tipos de auditoría	26
Tabla 2 Áreas de auditoría	27
Tabla 3 Proceso de Auditoría	29
Tabla 4 Definición de tipos de calidad de datos	35
Tabla 5 Características a auditar.....	36
Tabla 6 Tipos de procesamiento	38
Tabla 7 Preguntas de investigación	39
Tabla 8 Relevancia de documentos por base bibliográfica	42
Tabla 9 Selección de artículos	43
Tabla 10 Tipos de estadística.....	49
Tabla 11 Tipos de análisis.....	50
Tabla 12 Tecnologías de análisis avanzado de datos.....	53
Tabla 13 Técnicas Descriptivas.....	58
Tabla 14 Fases de metodología CRISP-DM	61
Tabla 15 Estructura de reporte de modificación de notas	65
Tabla 16 Evaluación de clúster ing.....	80
Tabla 17 Recursos de auditoría	82
Tabla 18 Descripción de Costos.....	84
Tabla 19 Riesgos de auditoría.....	85
Tabla 20 Plan de trabajo	87
Tabla 21 Preguntas de encuesta.....	90
Tabla 22 Preguntas de entrevista.....	92
Tabla 23 Tabla Hallazgos.....	115
Tabla 24 Preguntas de evaluación de calidad.....	122
Tabla 25 Resultados de Calidad	126
Tabla 26 Preguntas de evaluación de seguridad	126
Tabla 27 Resultados de seguridad.....	128

Tabla 28 Preguntas de evaluación de cumplimiento.....	128
Tabla 29 Resultados de cumplimiento	133

Índice de Figuras

Figura 1 Árbol de problemas	18
Figura 2 Alcance de proyecto	20
Figura 3 Metodología de proyecto	21
Figura 4 Etapas de auditoría de datos	33
Figura 5 Calidad de datos	35
Figura 6 Diagrama de búsqueda.....	41
Figura 7 Relación de inteligencia artificial, aprendizaje automático y aprendizaje profundo.	51
Figura 8 Relación de técnicas avanzadas	52
Figura 9 Técnicas de minería de datos	57
Figura 10 Cuadrante Mágico de Gartner	59
Figura 11 Metodología CRISP-DM	60
Figura 12 Módulos UTN	64
Figura 13 Curva elbow	68
Figura 14 Clúster 1 Nota1 vs Nota2.....	69
Figura 15 Clúster 2 Not_1 vs Not_2.....	70
Figura 16 Clúster 3 Not_3 vs Not_4.....	71
Figura 17 Clúster 4 Not_5 vs Not_6.....	72
Figura 18 Clúster 5 Not_7 vs Not_8.....	73
Figura 19 Clúster 6 Not_9 vs Not_10.....	74
Figura 20 Clúster 7 Not_11 vs Not_12.....	75
Figura 21 Clúster 8 Not_13 vs Not_14.....	76
Figura 22 Clúster 9 Not_15 vs Not_16.....	77
Figura 23 Clúster 10 Not_17 vs Not_18.....	78
Figura 24 Clúster 11 Not_19 vs Not_20.....	79
Figura 25 Pregunta 1	95
Figura 26 Pregunta 2	95

Figura 27 Pregunta 3	95
Figura 28 Pregunta 4	96
Figura 29 Pregunta 5	97
Figura 30 Pregunta 6	97
Figura 31 Pregunta 7	98
Figura 32 Pregunta 8	98
Figura 33 Pregunta 9	99
Figura 34 Pregunta 10	99
Figura 35 Pregunta 11	100
Figura 36 Pregunta 12	100
Figura 37 Pregunta 13	101
Figura 38 Pregunta 14	101
Figura 39 Pregunta 15	102
Figura 40 Pregunta 16	102
Figura 41 Pregunta 17	103
Figura 42 Pregunta 18	103
Figura 43 Pregunta 19	104
Figura 44 Pregunta 20	104
Figura 45 Pregunta 21	105
Figura 46 Precisión de datos FACAE	110
Figura 47 Precisión de datos FCCSS	111
Figura 48 Precisión de datos FECYT	111
Figura 49 Precisión de datos FICA	112
Figura 50 Precisión de datos FICAYA	113
Figura 51 Precisión de datos UTN	113
Figura 52 Dispersión de Fechas	114

Resumen

La implementación de sistemas académicos que facilita la gestión de datos e información relacionada con el personal docente, administrativo y estudiantil de las instituciones de educación superior, provoca una creciente cantidad de información, dificultando el análisis de la información dentro del proceso de auditoría informática.

El objetivo de este estudio es ayudar a los auditores a realizar procesos de auditoría informática más eficientes utilizando técnicas de procesamiento y análisis de datos avanzadas. Se efectuó una revisión exhaustiva de la literatura para establecer una base teórica sólida. Se seleccionaron tres técnicas de minería de datos: agrupación, asociación y correlación, para identificar patrones significativos en grandes conjuntos de datos. En el caso de estudio se empleó la técnica de agrupación, se centró en el módulo de notas y calificaciones del sistema integrado universitario (SIIU) de la Universidad Técnica del Norte. Para este propósito, se utilizó la metodología de minería de datos CRISP-DM.

El enfoque de la auditoría informática se centró en el historial de modificación de notas para evaluar la calidad de los datos del sistema académico. Se verificó la confidencialidad, precisión, trazabilidad y comprensibilidad de acuerdo con la norma ISO/IEC 25012:2008. Se obtuvieron resultados que revelaron niveles variables de imprecisión en diferentes facultades, con un rango del 5% al 30% de datos imprecisos.

Este estudio demostró cómo la aplicación de técnicas avanzadas de procesamiento y análisis de datos puede mejorar los procesos de auditoría informática y ayudar a detectar errores en los datos del sistema académico.

Palabras clave: Auditoría informática, minería de datos, big data, agrupación, correlación, asociación, crisp-dm

Abstract

The implementation of academic systems that facilitate the management of data and information related to faculty, staff, and students of higher education institutions, causes an increasing amount of information, making it difficult to analyze the information within the computer audit process.

The objective of this study is to help auditors perform more efficient computer audit processes using advanced data processing and analysis techniques. A comprehensive literature review was conducted to establish a sound theoretical basis. Three data mining techniques, clustering, association, and correlation were selected to identify meaningful patterns in large data sets. The clustering technique was used in the case study, focused on the grades and grades module of the integrated university system (SIU) of the Universidad Técnica del Norte. For this purpose, the CRISP-DM data mining methodology was used.

The focus of the computer audit was on the grade modification history to evaluate the quality of the data in the academic system. Confidentiality, accuracy, traceability, and comprehensibility were verified in accordance with ISO/IEC 25012:2008. Results were obtained that revealed varying levels of inaccuracy in different faculties, with a range of 5% to 30% of inaccurate data.

This study demonstrated how the application of advanced data processing and analysis techniques can improve computer audit processes and help detect errors in academic system data.

Keywords: Computer auditing, data mining, big data, clustering, correlation, association, crisp-dm

Introducción

Antecedentes

El incremento del 27,81% al 31,21% de la tasa bruta de matrícula en Instituciones de Educación Superior (IES) en Universidades y Escuelas Politécnicas al 2021 (INEC, 2021) y la implementación de sistemas académicos, que facilitan la gestión de notas, tareas, entre otros, derivan en la generación de una gran cantidad de datos e información que complican la tarea del auditor.

Con la aplicación de técnicas de big data y el análisis de datos es posible realizar pruebas de auditoría sobre un porcentaje muy alto o incluso la totalidad de una población. En la actualidad existen técnicas que ya se aplican, y que están relacionadas con la inteligencia artificial o aprendizaje automático (Blázquez Hernández, 2018). Con base en a la información analizada, sin embargo, aunque, no se especifica una técnica avanzada de datos, que ayude a mejorar la productividad en el proceso de auditoría informática en el módulo del sistema académico de la UTN.

Situación actual

Actualmente, los sistemas académicos de las IES con mayor cantidad de recursos han logrado automatizar los procesos de horarios, notas, aulas, laboratorios, entre otros, siendo beneficioso para la comunidad académica (Cevallos, Cuesta, & Montero, 2018). La gran cantidad de datos generados por los sistemas académicos, sumando a la virtualización de las clases provocada por la pandemia, hace complicado el análisis de información en el proceso de auditoría.

En los últimos años, hubo un creciente interés en el uso de modelos de big data para apoyar las funcionalidades avanzadas de análisis de datos. Muchas empresas y organizaciones carecen de experiencia en tecnologías de la información (TI) y presupuesto

adecuado para beneficiarse de estas tecnologías (Redavid, Corizzo, & Malerba, 2018). Estas técnicas pueden ayudar a mejorar la productividad en el proceso de auditoría informática en el módulo del sistema académico de la UTN, reduciendo una de las grandes dificultades que tienen que afrontar los auditores al tener que manipular grandes volúmenes de información, y que por ende provoca auditorías subjetivas y fuera del tiempo acordado.

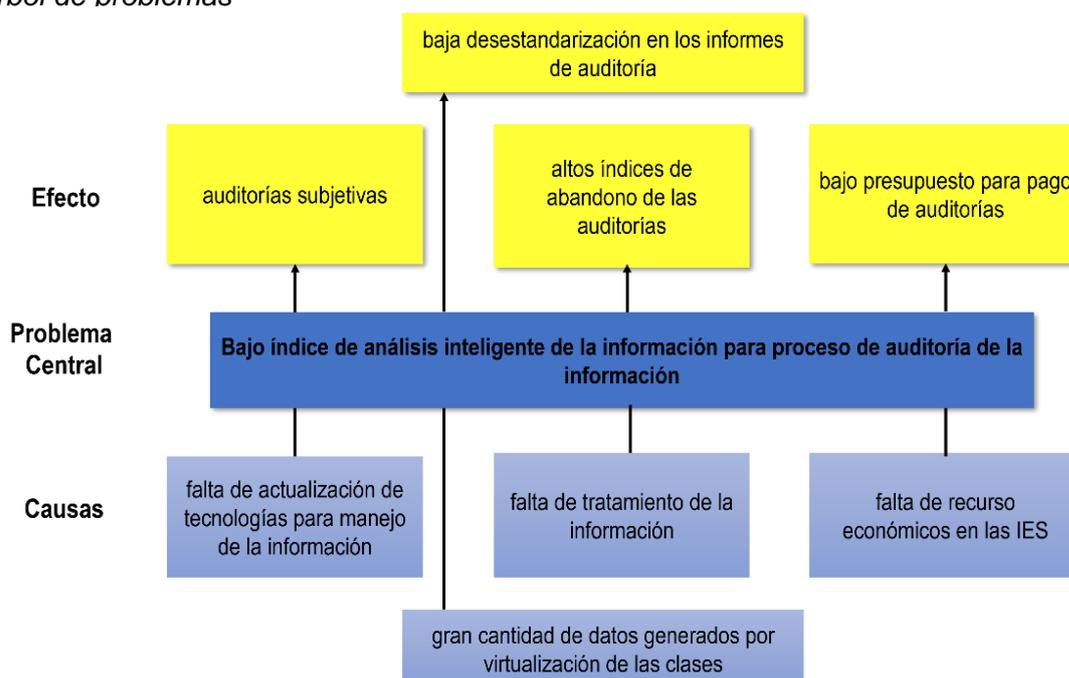
Planteamiento del Problema

Al implementar varias tecnologías para la automatización de procesos de un sistema de información académica y así poder satisfacer a los usuarios, generan gran cantidad de información, como consecuencia de esto se forman cuellos de botella que dificultan el análisis de la información dentro del proceso de auditoría informática en el módulo del sistema académico de la UTN (Astrid Shofi Dzihni, 2019).

A continuación, en la Figura 1, se identifica los problemas, causa y efectos que conlleva el bajo índice de análisis inteligente de la información para proceso de auditoría de la información.

Figura 1

Árbol de problemas



Objetivos

Objetivo General

Estudiar técnicas avanzadas de análisis de datos para el proceso de auditoría informática en el módulo del sistema académico de la Universidad Técnica del Norte (UTN).

Objetivos Específicos

- Realizar un marco teórico sobre tres técnicas para el procesamiento de datos.
- Utilizar técnicas de procesamiento de datos para aplicar proceso de auditoría informática en la carrera de software de la UTN.
- Validar resultados obtenidos dentro del proceso de auditoría informática.

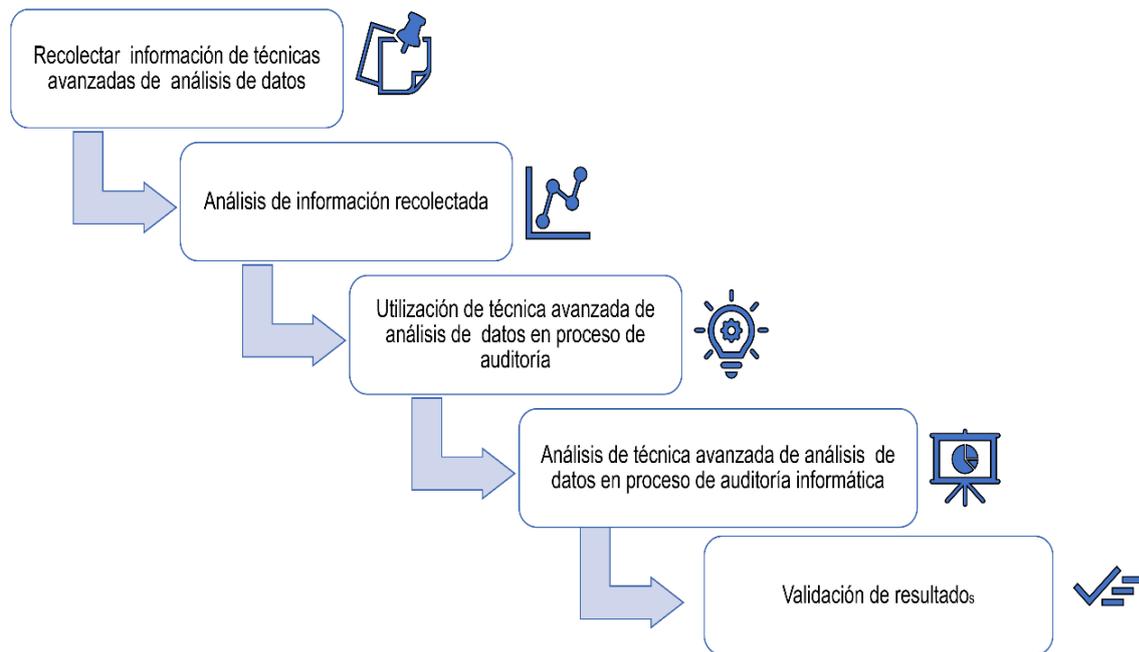
Alcance y Metodología

Alcance

El proyecto tiene como finalidad realizar un estudio basado en tres técnicas avanzadas de análisis de datos para innovar el proceso de auditoría de la información, se tomará las técnicas de procesamiento de datos más utilizadas de los últimos años, con el cuadrante mágico de Gartner y de esta manera poder utilizar una de las tres técnicas estudiadas, empleando los datos del módulo del sistema académico de la carrera de software del periodo 2020-2021 del segundo ciclo (Marzo – Agosto) en la UTN que se encuentren en su registro de base de datos. Y por consiguiente poder usar la técnica de procesamiento de datos, aplicando a un proceso de auditoría informática que se lo define en los siguientes pasos como, planificación, recopilación de información, muestreo de información, ejecución, reporte, acción correctiva y seguimiento. De esta manera se validarán los resultados obtenidos, los cuales se verán reflejados mediante un informe técnico, que permitirá al auditor verificar si es posible optimizar su tiempo, mejorando la calidad de la auditoría informática.

Figura 2

Alcance de proyecto



Metodología

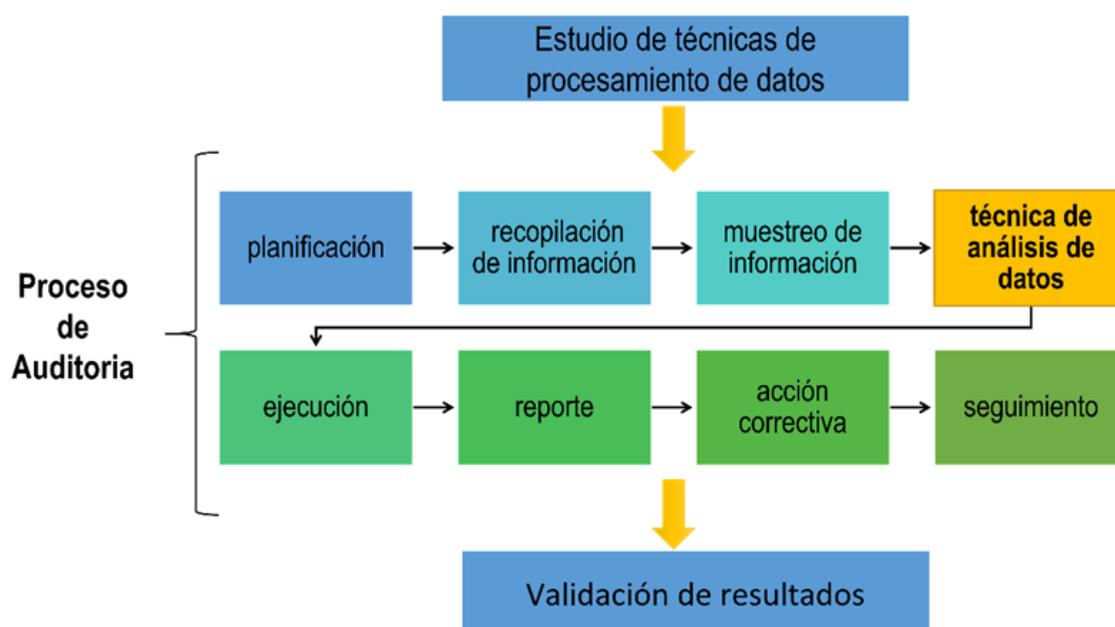
Para alcanzar el primer objetivo se realizó una investigación en diferentes fuentes bibliográficas, identificando tres técnicas avanzadas de procesamiento de datos mediante el cuadrante mágico de Gartner.

Posteriormente, para el segundo objetivo se efectuó un ejercicio de auditoría informática utilizando una técnica de análisis de datos con datos obtenidos del módulo del sistema académico de la carrera de software en la UTN.

Por último, en el tercer objetivo se validó los datos obtenidos con una encuesta que se destinará a expertos en el área de auditoría.

Figura 3

Metodología de proyecto



Justificación y Riesgos

Justificación

El presente proyecto está enfocado en dos objetivos del desarrollo sostenible (ODS) y en los planes de gobierno, descritos a continuación:

El objetivo 9: Industria, innovación e infraestructuras, el cual promueve construir nuevas infraestructuras resilientes, fomentar la innovación tecnológica (Naciones Unidas, 2021), ideal para la innovación de nuevas técnicas de análisis de datos en los procesos de auditoría y la cual permite mantener a las IES con infraestructura acorde a sus necesidades, sin embargo, la meta 9.4 del objetivo 9 se aplica a este proyecto por la adaptación de nuevas tecnologías que buscan establecer un análisis de datos para el proceso de auditoría informática en el módulo del sistema académico de la Universidad Técnica del Norte.

Objetivo 16: Paz, justicia e instituciones sólidas, el cual promueve sociedades justas, pacíficas e inclusivas (Naciones Unidas, 2021), ideal para mantener instituciones transparentes, aplicando la justicia y la paz laboral, la meta 16.6 del objetivo 16 se aplica a

este proyecto, ya que se busca mantener la transparencia al momento de manipular datos dentro de los IES, de manera similar la meta 16.10 también se aplica a este proyecto porque busca proteger la información con las leyes del estado y en este caso de estudio también obedeciendo la ley Orgánica de Educación Superior.

Plan de gobierno Objetivo 5: Impulsar la productividad y competitividad para el crecimiento económico sostenible de manera redistributiva y solidaria, el presente proyecto se base en punto 5.6 que promueve el desarrollo de tecnología transparente y la protección de la propiedad intelectual (Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe, 2021).

Plan de gobierno Objetivo 8: Promover la transparencia y la corresponsabilidad para una nueva ética social (Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe, 2021), dentro del cual el punto 8.2, referente a la transparencia de gestión de instituciones públicas y privadas, es ideal para la realización de este proyecto, ya que se basa en innovar procesos transparentes de auditoría.

Justificación Tecnológica.

Las técnicas avanzadas de procesamiento de datos se aplican intensamente en muchas instituciones para recopilar, transformar y analizar enormes conjuntos de datos para identificar patrones, apoyar argumentos y tomar decisiones. La aplicación de estas técnicas en la auditoría puede acelerar el proceso de prácticas de auditoría (Moutaz Haddara, 2018).

La aplicación de estas técnicas en la auditoría, junto con otros desarrollos tecnológicos, están despertando un gran interés tanto en la investigación como en la práctica. Junto con otros desarrollos tecnológicos (Felix Krieger, 2018). Por tal motivo, el presente proyecto tiene la intención de analizar técnicas avanzadas de procesamiento de datos aplicables al proceso de auditoría de la información y así reducir el tiempo de la auditoría y ayudando al auditor a obtener datos más precisos.

Justificación Metodológica.

El proyecto está basado en la investigación aplicada, que busca evaluar los resultados y experiencias del estudio de técnicas de procesamiento de datos en un grupo de datos, con

la intención de innovar los procesos de auditoría de la información (Merilampi, Ihanakangas, & Virkki, 2019), con el objetivo de recopilar información y seleccionarla posteriormente para el uso respectivo durante el proceso de desarrollo del proyecto.

CAPÍTULO I

Marco Teórico

1.1. Introducción

Debido al bajo índice de análisis inteligente de la información en proceso de auditoría, se busca innovar los procesos de auditoría informática, utilizando analítica avanzada de datos para agilizar tareas y ahorrar recursos examinando grandes cantidades de datos. Se realizó la revisión de la literatura para formar una base teórica robusta, recopilando varios artículos referentes al tema de técnicas avanzadas de procesamiento y análisis de datos.

En el estudio, se eligieron las tres técnicas de minería de datos para la detección de patrones en grandes conjuntos de datos como son, agrupación, asociación y correlación, que ayudaron al auditor a encontrar hallazgos relevantes en la auditoría. En este caso de estudio se empleó técnica de agrupamiento por brindar resultados fáciles de analizar, cuando se aplicó al conjunto de datos del módulo de notas y calificaciones del sistema integrado universitario (SIIU) de la Universidad Técnica del Norte, se utilizó la metodología de minería de datos CRISP-DM para aprovechar al máximo los datos en el procesamiento y análisis de estos.

La realización del proceso de auditoría informática se centró en el historial de modificación de notas para comprobar la calidad de los datos del sistema académico, basándonos en la norma ISO/IEC 25012:2008 verificando la confidencialidad, precisión, trazabilidad y comprensibilidad, en donde únicamente se encontraron hallazgos de imprecisión de 11% los datos en la facultad FACAE, 10% de imprecisión en la facultad FICAYA, 12% de imprecisión en la facultad FICA, 5% de imprecisión en la facultad FECYT y 30% de imprecisión en la facultad FCCSS.

1.2. Señalamiento de Variables

- **Variable dependiente:** - Análisis de datos orientados a auditoría

- **Variable independiente:** - Sistemas académicos
- Instituciones de educación superior
- Auditoría Informática

1.3. Auditoría

Según (Manrique Plácido, 2019) “La auditoría es un proceso sistemático para obtener y evaluar evidencias de manera objetiva y se aplica en distintas actividades de la organización social: empresas privadas y públicas, entidades de otros sectores, ámbito fiscal, operacional, medioambiental, forense, informático, entre otras. Así podemos distinguir los diferentes tipos de auditoría que existen”.

Las Auditorías tienen como objetivo general validar el cumplimiento de políticas y procedimientos; analizar que las actividades que se ejecutan permitan el desarrollo adecuado de los procesos (Arciniega Vera & Ludeña Gonzalez, 2014). La auditoría es un ejercicio significativo para la toma de decisiones dentro de la empresa, ya que permite descubrir falencias, prevenir errores y realizar cambios que ayuden a la institución a superarse y a mantenerse en el mercado (Zambrno Montesdeoca, 2020).

1.4. Tipos de Auditoría

Existen varios tipos de auditorías dependiendo de la industria o actividad que realiza la organización; Sin embargo, una característica que siempre comparten es que la información que obtienen en este ejercicio es de vital importancia para que una empresa o institución se desarrolle de manera eficiente (Paredes Murcia & León Cárdenas, 2021). Los tipos de Auditoría se pueden clasificar por el sujeto que lo efectúa, por su contenido y fines, por su amplitud y por su frecuencia, como se describe en la Tabla 1.

Tabla 1*Tipos de auditoría*

Por el Sujeto que la Efectúa	
Interna	Externa
Por su Contenido y Fines	
Auditoría de gestión	Auditoría organizativa
Auditoría operacional	Auditoría financiera
Auditoría contable	Auditoría informática
Por su Amplitud	
Auditoría total	Auditoría parcial
Por su Frecuencia	
Auditoría permanente	Auditoría ocasional

1.5. Auditoría Informática

El proceso de auditoría informática permite la recopilación, agrupación y evaluación de evidencia para determinar si un sistema de información cumple de manera correcta con su funcionalidad, conservando la accesibilidad, integridad y disponibilidad de los datos. Lo que consiente en verificar la adecuada utilización y operatividad de los recursos para que la organización cumpla con los fines propuestos. De esta forma, la auditoría informática sustenta y confirma la consecución de los objetivos tradicionales de auditoría (González Mataix, 2018).

Según (Rivas, 1988) en su libro fundamenta el contexto actual de la auditoría informática como “el conjunto de técnicas, actividades y procedimientos, diseñado para el análisis, evaluación, verificación y asesoramiento en asuntos relacionados con el desarrollo, la supervisión, la eficacia, la seguridad y la idoneidad de su servicio de información de la organización. El proceso de detección de fraudes o fallos en Tecnologías de la información es conocido como auditoría informática”.

1.5.1. Áreas de la Auditoría Informática

La auditoría es un órgano de control en algunas instituciones privadas y públicas y con base en la revisión bibliográfica, la auditoría Informática se divide en diferentes áreas como las que se marcan a continuación en la Tabla 2.

Tabla 2

Áreas de auditoría

Áreas de Auditoría	Definición
Auditoría de la gestión informática	La auditoría de gestión de la información es una colección de técnicas, actividades y procedimientos para recomendar en asuntos relativos a la planificación, control y adecuación del servicio informático en la empresa (Ponce Rodriguez, 2016).
Auditoría de sistemas	Auditoría de sistemas es la que se encarga de llevar a cabo la evaluación de normas, controles, técnicas y procedimientos que se tienen establecidos en una empresa para lograr confiabilidad, oportunidad, seguridad y confidencialidad de la información que se procesa a través de computadores (Tamavo Alzate, 2001).
Auditoría de los datos	La auditoría de datos es la evaluación de la calidad o utilidad de los datos para una intención en específica. La auditoría de datos, a diferencia de la auditoría de finanzas, implica analizar métricas clave, distintas de la cantidad, para crear conclusiones sobre las propiedades de un conjunto de datos (Garavit, 2020).

Auditoría de las bases de datos

Una auditoría a Base de Datos consiste en realizar un seguimiento constante y detallado de los controles establecidos a los sistemas de BD para garantizar la seguridad y el correcto uso de los datos almacenados por los usuarios (Villalobos Murillo, 2006) (Beltran Camacho & Bautista Cardenas, 2019).

Auditoría de la seguridad.

Una auditoría de seguridad de la información analiza cómo se protege la seguridad de la información de una institución, cubre temas que van desde la auditoría de la seguridad física de los centros de datos hasta la auditoría de la seguridad lógica de las bases de datos y destaca los componentes clave que se deben buscar y los diferentes métodos para auditar estas áreas (GHORPADE & MORE, 2015).

1.6. Proceso de Auditoría Informática

La duración de una auditoría varía según su alcance; las auditorías de alcance amplio pueden demorar varios meses, mientras que las auditorías de alcance limitado pueden demorar solo una semana o dos (The University of Texas at Austin, 2023).

La mayoría de las metodologías, marcos, estándares y procesos autorizados de auditoría informática comparten una estructura común de fases: planificación, trabajo de campo e informes, a pesar de que no existe un proceso estándar aceptado que se pueda aplicar en todos los contextos (Gantz, 2014).

En la tesis titulada "Propuesta de estandarización del proceso de auditoría de información para las instituciones de educación superior de la zona 1 del Ecuador mediante

normativas internacionales” se define el proceso a seguir para la realización de una auditoría informática, que se divide en cinco fases, Planificación, Ejecución, Comunicación de resultados, Validación y Seguimiento (Imbaquingo & Díaz, 2023), las cuales se detallan en la Tabla 3.

Tabla 3

Proceso de Auditoría

PROCESO DE AUDITORÍA (MAIIES)
PLANIFICACIÓN (FASE I)
<ul style="list-style-type: none"> ▪ Contacto con el cliente ▪ Estudio del entorno a auditar <ul style="list-style-type: none"> ▪ Equipo de auditoría ▪ Elaboración del Plan Preliminar ▪ Elaboración de propuesta de auditoría <ul style="list-style-type: none"> ▪ Contrato
EJECUCIÓN (FASE II)
<ul style="list-style-type: none"> ▪ Elaboración plan de trabajo ▪ Elaboración de instrumentos para la investigación de campo <ul style="list-style-type: none"> ▪ Análisis de la información recolectada ▪ Elaboración de las conclusiones preliminares para pasarlas a discusión
COMUNICACIÓN DE RESULTADOS (FASE III)
<ul style="list-style-type: none"> ▪ Elaboración del informe final preliminar <ul style="list-style-type: none"> ▪ Lectura del borrador ▪ Entrega del informe final
VALIDACIÓN (FASE IV)
<ul style="list-style-type: none"> ▪ Calidad

- Seguridad de la información
 - Cumplimiento
 - Análisis de datos

SEGUIMIENTO (FASE V)

- Contacto con el cliente
- Informe final de auditoría

Nota: Fuente (Imbaquingo & Díaz, 2023)

1.6.1. Fase I: Planificación

En la fase de planificación es una de las más importantes debido a que se realiza un cronograma a seguir con todas las actividades a realizar durante la ejecución de la auditoría, donde determinan los alcances, técnicas a utilizar, herramientas de auditoría, la definición de las pruebas y de los instrumentos para la recolección y procesamiento de información (papeles de trabajo o evidencia documental) (Imbaquingo & Díaz, 2023).

En esta fase también se procede a un rápido ejercicio de conocimiento del negocio y el sector al cual pertenece el contratante o cliente, para poder estimar y dimensionar el volumen de trabajo, identificar las particularidades y generalidades, prever las posibles dificultades y presupuestar los recursos humanos, físicos, financieros, de información y de tiempo (Sánchez Ch , 2021). En esta fase se realizan procedimientos como:

- Identificación de los motivos para ejecutar la auditoría
- Realizar una visita previa al establecimiento que será auditado
- Establecimiento de los objetivos de la auditoría
- Definición de los puntos a evaluarse en la auditoría
- Generación de planes, temarios, planificación de presupuestos para la auditoría
- Reconocimiento y elección de métodos, técnicas, instrumentos y procedimientos necesarios para la auditoría
- Asignación de recursos y equipos para la auditoría

(Chiriboga Toapanta, 2022).

1.6.2. Fase II: Ejecución

La segunda etapa es la Fase de Ejecución de la Auditoría y es la ejecución de esta, y está determinada por las características propias, las actividades elegidas y los requerimientos estimados en la planeación (Gonzalez Mataix & Berná Martinez, 2018). En esta fase se espera obtener resultados relevantes o posibles incidencias que se pueden presentar en la organización (Imbaquingo & Díaz, 2023).

Esta fase consiste en el desarrollo del programa de auditoría diseñado y que contiene todos los algoritmos previstos, la aplicación de las pruebas y el manejo de los respectivos instrumentos, mediante los cuales se recopilará información para procesar y obtener la evidencia que irá a soportar el juicio para la emisión del respectivo informe (Sánchez Ch , 2021), se realizan acciones como:

- Ejecución de actividades planificadas para la auditoría
- Uso de herramientas e instrumentos para la auditoría
- Identificación y registro de las desviaciones o anomalías halladas
- Elaboración de las conclusiones preliminares para pasarlas a discusión
- Agrupación de la documentación generada durante la auditoría

(Chiriboga Toapanta, 2022).

1.6.3. Fase III: Comunicación de Resultados

En esta fase se realiza un informe preliminar donde se presenta los hallazgos, se da lectura al borrador y finalmente se elabora el informe final de auditoría en el cual se describen las situaciones que se han presentado, que contiene las conclusiones y recomendaciones correspondientes a los problemas y deficiencias encontrados durante la ejecución de la auditoría (González Mataix, 2018), cuenta con acciones como:

- Análisis de información y elaboración de informes con las novedades detectadas.
- Elaboración del informe de auditoría

- Exposición del informe de auditoría
(Chiriboga Toapanta, 2022).

1.6.4. Fase IV: Validación

En esta etapa se evalúa la eficacia de los controles internos y el cumplimiento, también se pueden realizar otras tareas requeridas para cumplir con los objetivos de la auditoría (The University of Texas at Austin, 2023), estas características pueden ser evaluadas por la persona auditada o por personas que no están involucradas en la auditoría, cuenta con acciones como:

- Evaluación de calidad.
- Evaluación de cumplimiento.
- Evaluación de seguridad de la información.

1.6.5. Fase V: Seguimiento

El proceso de auditoría informática conlleva un seguimiento a lo largo del tiempo de 6 a 12 meses después de que se haya finalizado el informe final de auditoría, en el cual se realiza el contacto con el cliente para verificar si las recomendaciones fueron aplicadas con éxito y así garantizar el correcto funcionamiento del sistema y la calidad de los datos (Imbaquingo & Díaz, 2023).

1.7. Auditoría de Datos

En el proceso de auditoría de datos, es necesario evaluar si los datos cumplen con los objetivos. Este proceso implica perfilar el desempeño y descubrir el impacto de los datos de mala calidad en el desempeño de la organización. Y es evaluar la utilidad o calidad para fines particulares, los datos implican analizar medidas clave para crear conclusiones sobre la propiedad del conjunto de datos. Y la auditoría de datos es para verificar si la información utilizada es de naturaleza cualitativa.

Como menciona (Pacheco Curi, 2018, pág. 12) en su libro. “La dificultad de en el manejo de grandes volúmenes de información, la necesidad de disponer de

información íntegra, oportuna, segura y confiable, dio origen a la revolución informática, la cual ha generado una creciente dependencia para las empresas y usuarios en general, que se benefician diariamente de ella con el registro y procesamiento de operaciones; por consiguiente surge la imperiosa necesidad de ejercer control en este y es a través de la Auditoría de Datos la cual nos permite analizar la calidad de los datos”. La mala calidad de los datos presenta un enorme riesgo para la organización.

Datos incorrectos no solo lo expone a posibles fallos, también contribuye a la mala toma de decisiones en las operaciones de la organización. La auditoría de calidad de datos proporciona una vista completa del estado de los datos del sistema para prevenir y corregir posibles fallos (Jones, Ross, & Ruusalepp, 2009), los registros pueden servir como principal herramienta para una auditoría sistemática de la calidad del sistema (Planas, Rodríguez , & Lecha, 2022).

1.7.1. Etapas de Auditorías de Datos

La auditoría de datos está conformada por 4 etapas, estas están diseñadas para ejecutarse secuencialmente (Jones, Ross, & Ruusalepp, 2009), como se muestra a continuación en la Figura 4.

Figura 4

Etapas de auditoría de datos



Nota: Fuente (Jones, Ross, & Ruusalepp, 2009)

Estas etapas están consideradas en las tres fases del proceso de auditoría informática que se detallan en la Tabla 3, la etapa de planificación entra en la fase I de planificación, la etapa de identificar y clasificar los activos de datos junto con la etapa de evaluar la gestión de los activos de datos entra en la fase II de ejecución y finalmente la etapa de informar los hallazgos y recomendar cambios está considerada en la fase III de dictamen de la auditoría.

1.7.2. Calidad de Datos

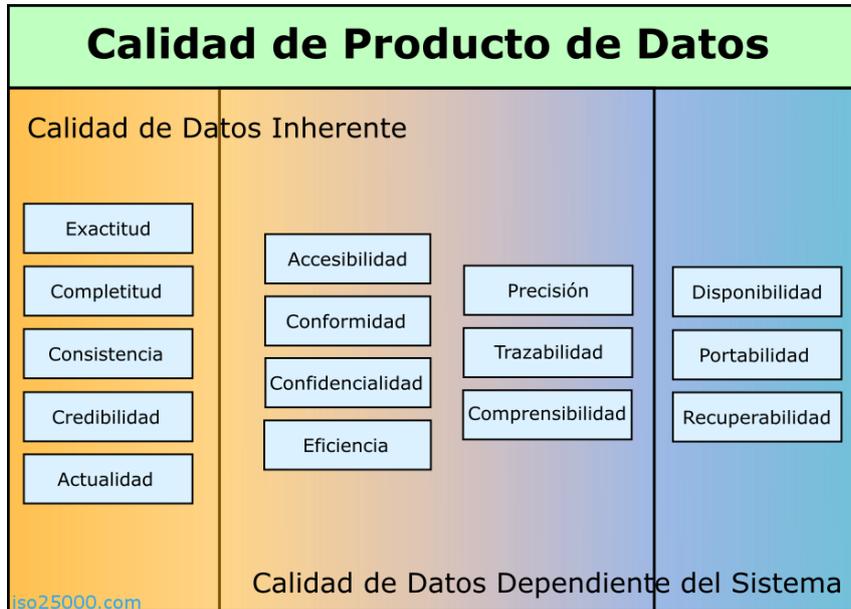
La evaluación del rendimiento del sistema depende en gran medida de la calidad de los datos, pero muchas organizaciones carecen de una estrategia para mejorar la calidad de los datos (Mustafa Ali, Anjum, & Kamel, 2018). El análisis de la calidad de los datos es esencial para el análisis del sistema, ya que afecta la precisión y la confiabilidad de los datos (Böttcher, Vieluf, Bruno, & al, 2022).

Evaluar la calidad de los datos de un sistema académico sirve para determinar si cumple con los objetivos, las leyes y los reglamentos de la institución, así como para tomar decisiones sobre el aprendizaje, el desempeño de los estudiantes y la seguridad de la información (Tejada Escobar, Murrieta Marcillo, & Villao-Santos, 2018).

El estudio del proyecto se basa en el uso de la norma ISO/IEC 25012:2008, sirve para evaluar la calidad de los datos de un producto de software, la cual cumple con 15 características tal como se muestra en la Figura 5.

Figura 5

Calidad de datos



Nota: Fuente: (ISO 25000, 2022)

La norma ISO/ IEC 25012:2008 se divide en tres conjuntos de características para evaluar la calidad de los datos como son datos inherentes, inherentes y dependientes del sistema y dependientes del sistema, los cuales se describen en la Tabla 4.

Tabla 4

Definición de tipos de calidad de datos

Calidad de Datos Inherente	Se refiere al grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas y necesarias cuando los datos son utilizados bajo condiciones específicas (ISO 25000, 2022).
Calidad de Datos Inherente y Dependiente del Sistema	Se refiere a las características cuya medición depende tanto de la naturaleza de los datos como de la definición del sistema de información en el que los datos se alojan.

Calidad de Datos	Se refiere al grado con el que la Calidad de Datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas (ISO 25000, 2022).
Dependiente del Sistema	

Se verifica el cumplimiento de las siguientes características que se muestra en la Tabla 5 en el proceso de auditoría informática para evaluar el reporte de modificación de notas para el período SEP2022-FEB2023, tomando como referencia la calidad de datos inherente y dependiente del sistema de la norma ISO/IEC 25012:2008.

Tabla 5

Características a auditar

Características	Definición
Confidencialidad	Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información (junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.
Precisión	Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.
Trazabilidad	Grado en el que los datos tienen atributos que proporcionan un camino de acceso

	auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico.
Comprensibilidad	Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.

Nota: Fuente (ISO 25000, 2022)

1.8. Procesamiento de datos

Según la real academia española define el procesamiento de datos como “aplicación sistemática de una serie de operaciones sobre un conjunto de datos, generalmente por medio de máquinas, para explotar la información que estos datos representan” (Real Academia Española , 2023). Que se entiende como la manipulación y transformación de datos en conocimientos significativos o información útil.

Las principales operaciones que se realiza en el procesamiento de datos es la limpieza de datos, integración, transformación, reducción, privacidad y protección de los datos (Sun, Cai, Li, & Liu, 2018).

1.8.1. Tipos de procesamiento de datos

Existen varios tipos de procesamiento de datos basado en el origen de los datos y las operaciones que se pretenden realizar para producir la salida, como se indica en la Tabla 6.

Tabla 6*Tipos de procesamiento*

Tipos	Usos
Procesamiento por lotes	Los datos se recopilan y procesan en lotes. Se utiliza para grandes cantidades de datos.
Procesamiento en tiempo real	Los datos se procesan en cuestión de segundos cuando se proporciona la entrada. Se utiliza para pequeñas cantidades de datos.
Procesamiento en línea	Los datos se introducen automáticamente en la CPU tan pronto como están disponibles. Se utiliza para el procesamiento continuo de datos.
Multiprocesamiento	Los datos se dividen en tramas y se procesan empleando dos o más CPU dentro de un solo sistema informático. También conocido como procesamiento paralelo.
Tiempo compartido	Asigna recursos informáticos y datos en intervalos de tiempo a varios usuarios simultáneamente.

Nota: Fuente (Duggal, 2023)

Los tipos de procesamientos más utilizados según diversos autores es procesamiento por lotes, en tiempo real.

1.8.2. Técnicas de procesamiento

Algunas técnicas avanzadas de procesamiento de datos más utilizadas en la actualidad incluyen el aprendizaje automático, el procesamiento del lenguaje natural, la minería de datos y la inteligencia artificial (Riquelme, Ruiz, & Gilbert, 2006). Estas técnicas

se utilizan en una amplia variedad de campos, desde la medicina hasta la industria manufacturera, y se han convertido en herramientas esenciales para el análisis y la toma de decisiones en el mundo empresarial y científico.

En el artículo publicado en la revista "IEEE Transactions on Knowledge and Data Engineering", en el que los autores describen un sistema de procesamiento de datos que utiliza técnicas de minería de datos para analizar grandes conjuntos de datos de sensores. El sistema utiliza algoritmos de clústering y de asociación para identificar patrones en los datos y extraer información valiosa. Los autores concluyen que su sistema puede ser utilizado en una amplia gama de aplicaciones, desde la monitorización de la salud hasta el mantenimiento predictivo de maquinaria (Plumed, Contreras Ochando, Ferri, & Hernández Orallo, 2021).

1.9. Revisión de la literatura

La revisión de la literatura es una etapa importante de cualquier trabajo de investigación ya que nos ayuda a posicionar el estudio y sustentarlo teóricamente y conceptualmente de lo que otros investigadores e investigadoras han escrito previamente sobre la temática (Arnau Sabatés & Sala Roca, 2020). Se puede realizar en cuatro fases: preguntas de investigación, búsqueda de documentos, selección de artículos y la extracción de datos relevantes (Webster & Watson, 2002), tal y como se detalla a continuación:

1.9.1. Preguntas de investigación

Las tres preguntas de investigación que se han planteado para el desarrollo del proyecto sirven como guía de revisión de este tema. En la Tabla 7 se describen las preguntas de investigación.

Tabla 7

Preguntas de investigación

N°	Preguntas de Investigación	Motivación
-----------	-----------------------------------	-------------------

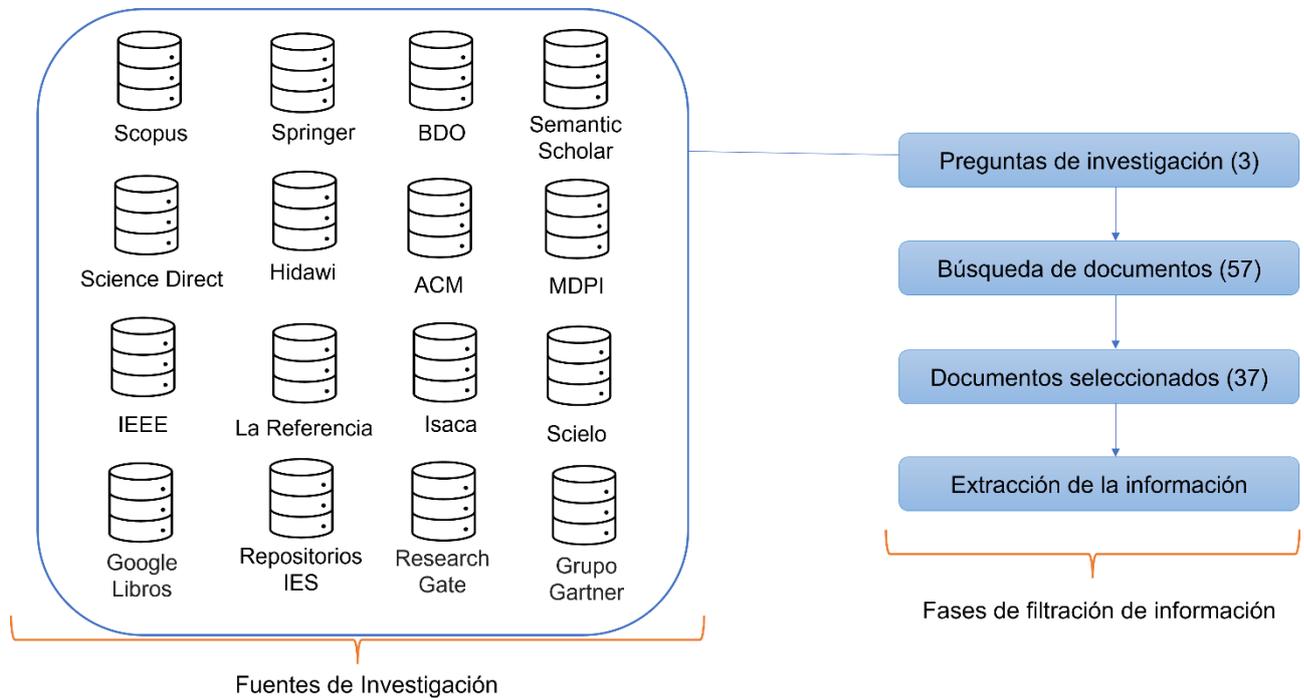
P1	¿Que son las técnicas de procesamiento y análisis de datos?	Conocer las características de las técnicas de procesamiento y análisis de datos
P2	¿Cuáles son las técnicas avanzadas de procesamiento y análisis de datos existentes?	Identificar las técnicas avanzadas que existen
P3	¿Qué técnicas avanzadas de procesamiento y análisis de datos se usan en auditoría?	Puntear las técnicas de procesamiento y análisis de datos que puedo utilizar en auditoría

1.9.2. Búsqueda de Documentos

Se consideran 16 bases de datos bibliográficas para la segunda fase, tomando en cuenta la cadena de búsqueda. (advanced AND data AND analysis AND techniques OR data AND Audit OR types AND of AND it AND audits), con un total de 57 documentos de diversas fuentes de investigación, los cuales pertenece 1 documento a ACM, BDO, Gartner, Hidawi, Semantic Scholar y Isaca, 2 documentos a MDPI, Scopus, 3 documentos a Google Libro, La Referencia y Science Direct, 5 a Scielo, 6 a Research Gate, Springer, 8 a IEEE y 13 a Repositorio IES. Los cuales se observan en la Figura 6 del diagrama de busqueda.

Figura 6

Diagrama de búsqueda



Selección de artículos

Se tomaron en cuenta tres pasos en la selección de artículos. Los criterios de inclusión y exclusión se aplican en la primera etapa. Los autores tomaron en cuenta los siguientes criterios de inclusión: (i) artículos científicos, (ii) revisiones de literatura y otros artículos indexados y no indexados, pero relevantes para la investigación.

En el segundo paso, se centró en la documentación más significativa de la revisión sistemática de la literatura (SLR). Esto se hizo al responder a tres preguntas de investigación: “(P1) ¿Que son las técnicas de procesamiento y análisis de datos?”, “(P2) ¿Cuáles son las técnicas avanzadas de procesamiento y análisis de datos existentes?” y “(P3) ¿Qué técnicas avanzadas de procesamiento y análisis de datos se usan en auditoría?”. El título, el resumen y las palabras clave son las primeras en ser analizadas.

Finalmente, en el tercer paso, se inspeccionan las secciones de introducción y conclusiones de la publicación para garantizar que la información contenida sea relevante

para la pregunta de investigación. El número total de documentos recuperados después de aplicar estas tres etapas se muestra en la Tabla 8.

Tabla 8

Relevancia de documentos por base bibliográfica

Base de datos bibliográficos	N° de Artículos	N° de Artículos relevantes
ACM	1	1
BDO	1	0
Grupo Gartner	1	1
Google Libro	3	3
Hidawi	1	1
IEEE	8	3
Isaca	1	1
La Refrencia	3	1
MDPI	2	2
Repositorio IES	13	7
Research Gate	6	4
Scielo	5	4
Science Direct	3	2
Scopus	2	0
Springer	6	6
Semantic Scholar	1	1

Total	57	37
--------------	-----------	-----------

En la Tabla 9 se muestra la selección de artículos de iteres realizada en las diferentes fuentes bibliográficas que se utilizaron para responder las preguntas de investigación.

Tabla 9

Selección de artículos

N°	Título	Autores	Base Bibliográfica	Año
1	Advanced query model design concept to support multi-dimensional data analytics for relational database management systems	(Liyanage, Nandika Habaraduwa)	IEEE	2017
2	Advanced data analysis techniques with marketing applications	(Paetz, Steiner, & Hruschka)	Springer	2022
3	Big data analytics: a survey	(Tsai, Lai, & Chao)	Springer	2015
4	Estadística descriptiva	(Rendón Macías, Villasís Keeve, & Miranda Novales)	Repositorio IES	2016
5	Componente de software de análisis prescriptivo basado en	(González Silva & Peña Rojas)	Repositorio IES	2022

	las variables oceánicas y meteorológicas de Colombia			
6	Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos	(Lemus Delgado & Pérez Navarro)	Repositorio IES	2019
7	Big Data Analytics Using Artificial Intelligence	(Gandomi, Chen, & Abualigah)	MDPI	2023
8	Exploring Data	(Mount & Zumel, 2016)	Google Libros	2016
9	Introducción a la estadística	(Ross, Sheldon)	Google Libros	2018
10	Statistical methods for the social sciences	(Agresti, Alan)	Google Libros	2018
11	Reflexiones sobre el uso de la estadística inferencial en investigación didáctica	(Veiga, Otero, & Torres)	Scielo	2020
12	Técnicas de minería de datos aplicadas en la detección de fraude: estado del arte	(Santamaria Ruiz, 2006)	Research Gate	2006
13	Magic Quadrant	(Gartner)	Grupo Gartner	2023
14	Clúster ing and visualizing the status of child health in kenya: a data mining approach	(Njiru & Opiyo)	Research Gate	2018

15	Inteligencia artificial y big data como soluciones frente a la covid-19	(Márquez Díaz)	Scielo	2020
16	Automated abnormality classification of chest radiographs using deep convolutional neural networks	(Tang, Tang, & Peng)	Springer	2020
17	Deep learning: a comprehensive overview on techniques, taxonomy, applications, and research directions	(Sarker)	Springer	2021
18	Generalidades del machine learning y su aplicación en la gestión sanitaria en servicios de urgencia	(Pedrero, Reynaldos Grandón, Ureta Achurra, & Cortez Pinto)	Scielo	2021
19	Big data	(Sha & Carotti Sha)	Springer	2016
20	EDISON Data science framework: a foundation for building data science profession for research and industry	(Demchenko, Belloum, & Los)	IEEE	2016

21	Audit data analysis and application based on correlation analysis algorithm	(Jifan & Muhammad)	Hidawi	2021
22	Técnicas de análisis de datos aplicaciones prácticas utilizando microsoft excel y weka	(Molina López & García Herrero)	Repositorio IES	2006
23	Application of data mining techniques for the analysis of academic efficiency	(Grijalva Arriaga, Freire Avilés, & Real Avilés)	La Referencia	2018
24	Detección de patrones de contrabando para la gestión de aprehensiones y retenciones, utilizando técnicas predictivas de clasificación y regresión de minería de datos	(Mancero Menoscal)	Repositorio IES	2020
25	Estudio de la seguridad en big data, privacidad y protección de datos mediante la iso/iec 27007:2017- aplicado a los datos académicos de la universidad técnica del norte	(Hernández Obando)	Repositorio IES	2019
26	Advanced data analytics for it auditors	(Alexiou)	ISACA	2017

27	Integración de los algoritmos de minería de datos 1r, prism e id3 a postgresql	(Robles Aranda & Sotolongo)	Scielo	2013
28	Application of machine learning algorithms in audit data analysis	(Zhou J.)	ACM	2022
29	Research on audit data analysis under the background of big data	(Zhang)	Research Gate	2021
30	Initial implementation of data analytics and audit process management	(Sanoran & Ruangprapun)	MDPI	2023
31	Design and implementation of university audit platform based on big data analysis	(Yang, Guohua, Wanlong, Qinghai, & Yixian)	Science Direct	2022
32	Using data mining as a tool for anomaly detection in food safety audit data	(Kleboth, Kosorus, Rechberger, & Luning)	Science Direct	2022
33	Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso	(Espino Timó)	Repositorio IES	2017
34	A detailed study of clustering algorithms	(Kamalpreet & Anuranjan)	IEEE	2017

35	Review on Mining Association Rule from Semantic Data	(Kale, 2016)	Semantic Scholar	2016
36	Feature selection based on mutual information with correlation coefficient	(Zhou, Wang, & Zhu)	Springer	2022
37	Clúster ing and Visualizing the Status of Child Health in Kenya: A Data Mining Approach	(Njiru & Opiyo)	Research Gate	2018

1.10. Técnicas de Procesamiento y Análisis de Datos

Las técnicas de procesamiento y análisis de datos son herramientas y métodos que se utiliza para recopilar, procesar y analizar información en orden para obtener ideas y conocimientos que son útiles para tomar decisiones, estas herramientas pueden abarcar estadística tradicional y ciencia de datos.

La ciencia de datos es un campo amplio que toca aspectos de la estadística, el aprendizaje automático, amplias áreas de inteligencia artificial, análisis de datos, reconocimiento de patrones, comprensión del lenguaje natural, y la manipulación masiva de datos (Gandomi, Chen, & Abualigah, 2023). La utilización de las herramientas, los métodos y el trabajo depende de manera considerable de varios aspectos como el dominio, problema y del punto de vista.

La dirección empresarial conduce a tomar nuevas estrategias, provocando a las instituciones o empresas que manejan sistemas pequeños y medianos a administrar bases de datos relacionales y modelos de consulta simples como técnica primaria de análisis de datos (Liyanage, 2017).

Sin embargo, esta diversidad puede parecer abrumadora y divisiva. Un estadístico tradicional puede no considerar el análisis de texto como ciencia de datos y, de manera

similar, alguien que usa redes neuronales para analizar imágenes puede no apreciar la inferencia estadística clásica. Por esto es importante saber definir una técnica de acuerdo con el problema que se presenta (Mount & Zumel, 2016).

La exploración y limpieza de datos son temas que a menudo se descuidan, pero también son las partes más cruciales y que consumen más tiempo del proceso de ciencia de datos. Sin buenos datos, no puede desarrollar modelos efectivos ni obtener información importante (Mount & Zumel, 2016).

1.10.1. Estadística Tradicional

“La estadística es el arte de aprender a partir de los datos. Está relacionada con la recopilación, su descripción subsiguiente y su análisis, lo que nos lleva a extraer conclusiones” (Ross, 2018, pág. 3).

La Real Academia Española lo define como “Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas Variables (cualitativas, cuantitativas) (Real Academia Española, 2023)”. La estadística se divide en descriptiva e inferencial tal como se observa en la Tabla 10.

Tabla 10

Tipos de estadística

Estadística Descriptiva	La estadística descriptiva es la rama de la estadística que hace recomendaciones sobre cómo analizar imágenes, tablas y figuras (Rendón Macías, Villasís Keeve, & Miranda Novales, 2016).
Estadística Inferencial	Las estadísticas inferenciales proporcionan predicciones sobre una población, basado en información de una muestra de la población (Agresti, 2018, pág. 17).

El análisis estadístico de datos inferencial proporciona herramientas que permiten la evaluación sistemática y eficaz de una muestra de la poblacional de estudio (Veiga, Otero, & Torres, 2020).

1.10.2. Ciencia de datos

La ciencia de datos tiene un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos (Bernal, 2023). A pesar de tener un camino novedoso, su amplia difusión y aceptación, el uso de técnicas basadas en la ciencia de datos en el campo de los estudios globales está comenzando (Lemus Delgado & Pérez Navarro, 2019).

Los resultados de la aplicación de la ciencia de datos ilustran que, con los métodos eficientes disponibles, podemos ser capaces de analizar los datos a gran escala en un tiempo razonable (Tsai, Lai, & Chao, 2015). En general, los análisis de datos se pueden clasificar en tres tipos de análisis, es decir, análisis descriptivos, análisis predictivos y análisis prescriptivos (Paetz, Steiner, & Hruschka, 2022), los cuales se detallan en la Tabla 11.

Tabla 11

Tipos de análisis

Análisis Descriptivos.	<p>“Los métodos de minería de datos y la agregación de datos son los principales temas de los análisis descriptivos. Las empresas e instituciones pueden revisar situaciones pasadas con la ayuda de este tipo de análisis.”^a</p> <p>Usa datos históricos para identificar el comportamiento. Es el análisis más utilizado y su propósito es obtener una instantánea importante de la situación para tomar decisiones altamente exitosas.</p>
-------------------------------	--

<p>Análisis Predictivos.</p>	<p>“El análisis predictivo ayuda a los gerentes a hacer predicciones sobre el futuro mediante el uso de métodos de pronóstico y modelos estadísticos.”^a</p> <p>Se basa en técnicas matemáticas avanzadas, como estadísticas o aprendizaje automático, para predecir los datos que faltan y describir lo que sucederá (UNIR Ingeniería y tecnología, 2020).</p>
<p>Análisis Prescriptivos</p>	<p>“Utilizando árboles de decisión y otras técnicas de optimización, el análisis prescriptivo incorpora los resultados del análisis predictivo para proponer proactivamente la correcta toma de decisión.”^a</p> <p>Este análisis emplea una variedad de técnicas de modelado y optimización. Sin embargo, las organizaciones que emplean este tipo de análisis con frecuencia se basan en procedimientos vitales esenciales para la productividad. (González Silva & Peña Rojas, 2022).</p>

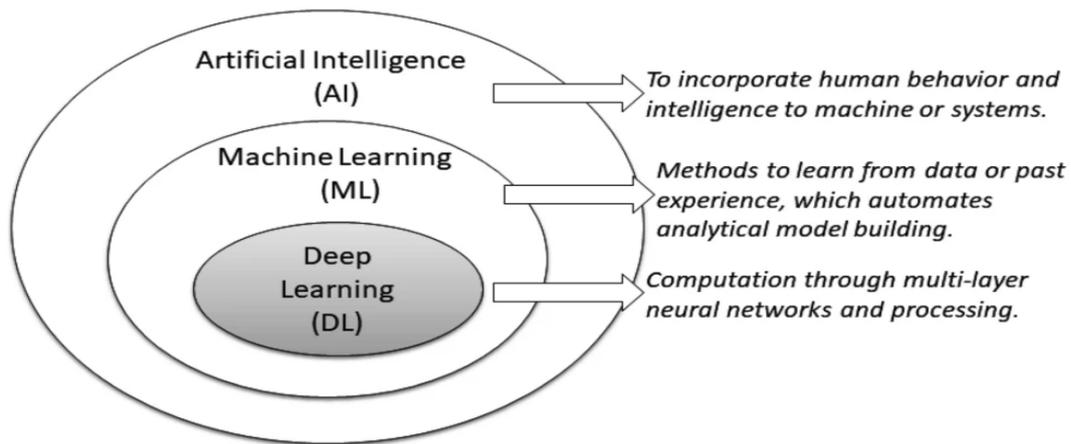
Nota: ^a (Paetz, Steiner, & Hruschka, 2022)

1.11. Técnicas Avanzadas de Procesamiento y Análisis de Datos

El procesamiento y análisis de datos avanzado se refiere a un examen de datos que emplea herramientas de datos sofisticadas y técnicas digitales que superan los paradigmas de análisis de datos tradicionales. En estos días, los términos inteligencia artificial (IA), aprendizaje automático (ML), y aprendizaje profundo (DL) se relacionan entre ellos (Sarker, 2021), como se observa en la Figura 7 y se utilizan con frecuencia para resolver diferentes problemas.

Figura 7

Relación de inteligencia artificial, aprendizaje automático y aprendizaje profundo.

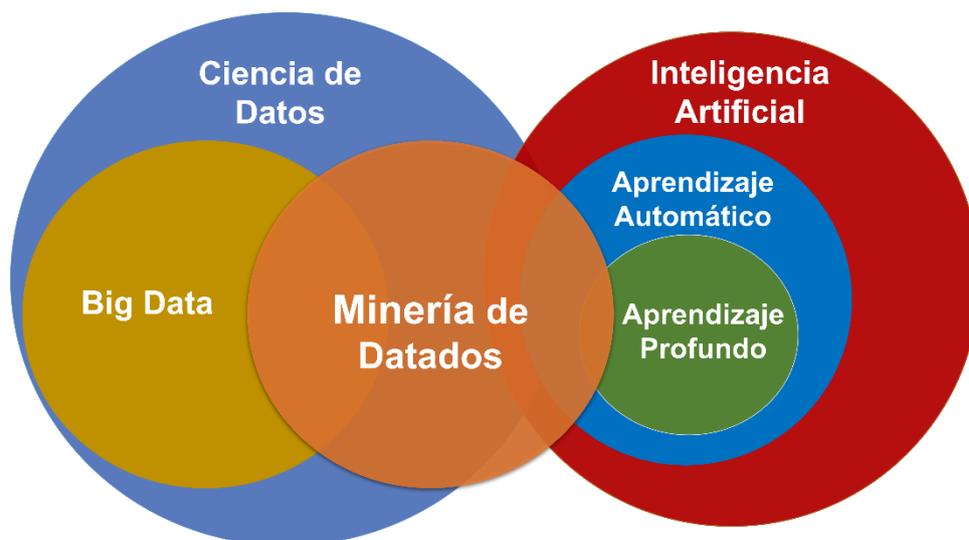


Nota. (Sarker, 2021, pág. 3)

El procesamiento y análisis avanzado de datos es un conjunto completo de técnicas y métodos analíticos como Big Data, Inteligencia Artificial, Machine Learning, Deep Learning, Data Science, Data Mining, como se observa en la Figura 8 que permiten un mejor análisis predictivo y brindan información sobre el cambio a medida que ocurre (Sdg Group, 2020), lo que permite a las organizaciones desarrollar mejores respuestas y actuar sobre la base de pronósticos y planes más precisos, permitiéndoles descubrir tendencias y patrones, resolver problemas, predecir el futuro e impulsar el cambio.

Figura 8

Relación de técnicas avanzadas



Cada tecnología de procesamiento y análisis avanzado de datos posee sus propias técnicas y algoritmo de acuerdo de lo que se pretenda hacer con base en los datos que se manejen, en la Tabla 12 se especifica la función de cada una de las tecnologías de análisis avanzado de datos.

Tabla 12

Tecnologías de análisis avanzado de datos

Ciencia de Datos	La ciencia de datos es una nueva área de estudio que solicita un enfoque multidisciplinario y debe construirse con una fuerte conexión con Big Data y tecnologías basadas en datos (Demchenko, Belloum, & Los, 2016).
Big Data	El big data es el campo que estudia la gestión y el procesamiento de cualquier conjunto de datos demasiado grande que puedan ser interpretados de forma continua y particular que cuenta con las características de volumen, velocidad y variedad (Sha & Carotti Sha, 2016).
Minería de Datos	La minería de datos utiliza algoritmos de regresión, clasificación y agrupamiento, entre otros, para analizar y segmentar enormes cantidades de datos y extraer patrones recurrentes, detectando tendencias y descubriendo hechos (Santamaria Ruiz, 2006).
Inteligencia Artificial	La inteligencia artificial abarca una variedad de campos, como el reconocimiento de voz, procesamiento del lenguaje natural, visión artificial, entre otros, donde el objetivo es que los sistemas permiten, razonar, participar, y aprender tareas relacionados a las de un humano (Márquez Díaz, 2020).

Aprendizaje Automático	El aprendizaje automático es una rama de la inteligencia artificial, cuyo propósito es el desarrollo de algoritmos para extraer información de los datos ya sea para explicación, clasificación o predicción de datos estructurados y datos no estructurados (Pedrero, Reynaldos Grandón, Ureta Achurra, & Cortez Pinto, 2021).
Aprendizaje Profundo	Otra área de especialización de la IA es el aprendizaje profundo, que se utiliza para evaluar la imagen, datos de vídeo y audio, utilizando redes neuronales convolucionales (CNN) con enormes conjuntos de datos etiquetados, para hacer diferentes tareas de visión por computador (Tang, Tang, & Peng, 2020).

1.12. Técnicas Avanzadas de Procesamiento y Análisis de Datos en Auditoría

Para la extracción del conocimiento de fuentes de datos masivas se emplean diferentes técnicas avanzadas de datos que utilizan algoritmos capaces de obtener relaciones entre distintos atributos o conceptos para ayudar a la toma de decisiones (Molina López & García Herrero , 2006). La alta deficiencia de los algoritmos tradicionales para análisis de datos de auditoría permite a las técnicas de minería de datos extraer evidencias ocultas de forma efectiva de cantidades intensivas de datos (Jifan & Muhammad, 2021).

La investigación realizada en el artículo (Jifan & Muhammad, 2021) indica que la minería de datos se puede utilizar como base para el desarrollo de modelos de análisis de auditorías avanzadas. El algoritmo de regla de asociación en el aprendizaje automático es el que utiliza en la minería de datos de auditoría, es actualmente una de las principales herramientas para resolver problemas de manera eficiente. (Zhou J. , 2022)

En el artículo (Alexiou, 2017) titulado “Análisis avanzado de datos para auditores de TI” se enumeran varias técnicas utilizadas en el área de auditoría, tomado en cuenta las tres

técnicas avanzadas como clasificación, agrupación y asociación. De igual manera se utilizan técnicas de Minería de Datos, árbol de decisión y Clústering para obtener patrones de comportamiento (Grijalva Arriaga, Freire Avilés, & Real Avilés , 2018).

En (Kleboth, Kosorus, Rechberger, & Luning, 2022) indica que la automatización para detectar anomalías la cual describe como una condición que está fuera del rango típico de variaciones o patrones en los datos de auditoría, se puede lograr mediante la aplicación de algoritmos adecuados como:

- basadas en frecuencias bayesianas.
- basadas en frecuencias condicionales.
- basadas en densidad.
- basadas en agrupamientos.
- basadas en distancias.
- teóricas de la información.
- basadas en compresión y semi supervisadas.

En el artículo (Yang, Guohua, Wanlong, Qinghai, & Yixian , 2022) explica el uso del big data en auditoría e indica que, el análisis de big data en auditoría analiza la tendencia de correlación de datos y se centra en la minería de datos, en contraste con el proceso de auditoría convencional basado en consultas, se deben usar los siguientes algoritmos para la auditoría de datos:

Los datos no estructurados se usan principalmente en minería de datos.

- agrupación.
- detección de eventos.
- aprendizaje inteligente.

Los datos estructurados son principalmente análisis estadísticos de datos.

En el artículo (Robles Aranda & Sotolongo, 2013) indica que las técnicas de minería de datos son un enfoque conceptual que con frecuencia se ponen en práctica mediante

diferentes algoritmos. Estos se pueden clasificar en técnicas de agrupación, asociación, predicción y clasificación en función de su utilidad.

El procesamiento y análisis de los datos de auditoría se está convirtiendo cada vez más crucial con el big data, tiene un impacto al realizar auditorías, convirtiéndola en una actividad de comportamiento constante (Zhang, 2021). Los auditores generalmente ven los beneficios de las herramientas de análisis de datos de auditoría como formas de administrar su proceso de auditoría de manera más efectiva (Sanoran & Ruangrapun , 2023).

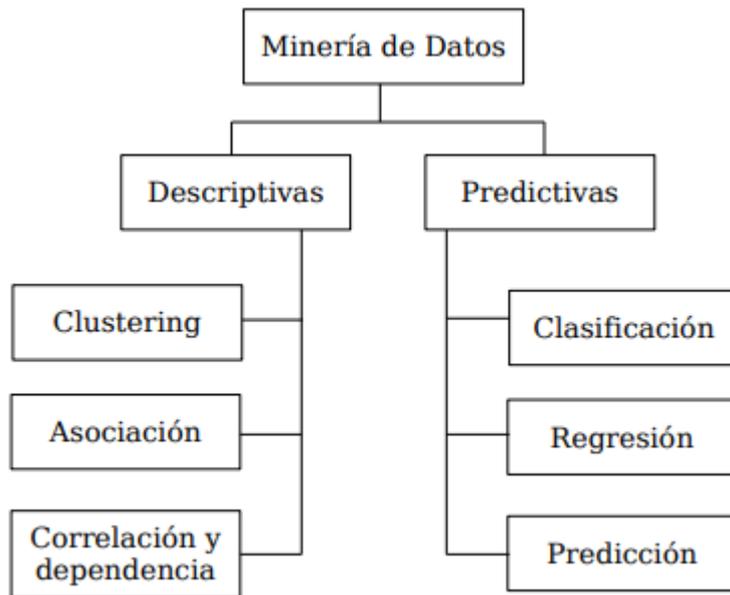
1.12.1. Selección de técnicas avanzadas

El procesamiento y análisis avanzado de datos se utiliza para obtener evidencia de auditoría, es la ciencia y el arte de encontrar y analizar patrones, desviaciones y discrepancias con el fin de extraer información útil más allá de los datos primarios sobre el tema auditado (Ávalos Andrade & Sainz-Grant , 2019).

La minería de datos por su descripción y la aplicación en varios trabajos relacionados con técnicas de procesamiento y análisis avanzado de datos en auditoría es seleccionada para la utilización del estudio. Las técnicas de minería de datos pueden dividirse en 2 categorías principales: predictivas (aprendizaje supervisado) y descriptivas (aprendizajes no supervisados) (Hernández Valadez, 2006) las cuales se describen en la Figura 9.

Figura 9

Técnicas de minería de datos



Nota: Fuente (Hernández Valadez, 2006)

Las técnicas predictivas se basan en conocimientos teóricos previos, especifican el modelo para los datos; el modelo debe cambiar después de la extracción de datos antes de ser validado (Mancero Menoscal, 2020) (Pérez López & Santin González, 2007), para estar en condiciones de anticipar acciones futuras.

Las técnicas descriptivas crean modelos que de alguna manera explican el dato; es decir, tienen como objetivo describir en lugar de predecir datos futuros (Mancero Menoscal, 2020) (Lara Torralbo, 2014).

Técnicas de Minería de Datos

De las diversas técnicas de minería de datos que se muestran en la Figura 9, limitamos la selección a las siguientes tres técnicas descriptivas que se muestran en la Tabla 13, basado en el objetivo del estudio de auditoría informática, que es identificar patrones en lugar de predecir datos para la toma de decisiones.

Los modelos descriptivos cuantifican las relaciones entre los datos de una manera que se usa con frecuencia para dividir a los clientes o contactos en grupos. Los modelos descriptivos reconocen varias relaciones, en contraste con los modelos predictivos, que se concentran en pronosticar el comportamiento de un usuario específico (Espino Timó, 2017).

Tabla 13

Técnicas Descriptivas

Técnicas	Descripción
Agrupación	Es una técnica de aprendizaje no supervisada que se utiliza para el análisis exploratorio de datos para descubrir patrones ocultos que existen en los datos que no se pueden distinguir con normalidad (Kamalpreet & Anuranjan , 2017).
Asociación	El aprendizaje de reglas de asociación es una técnica utilizada en la minería de datos para encontrar patrones intrigantes entre variables en grandes volúmenes de datos. Un número significativo de puntos de datos están conectados de maneras interesantes por la regla de asociación (Kale, 2016).
Correlación y dependencias	Estos métodos son ventajosos para encontrar conexiones y patrones ocultos en los datos. Con el fin de eliminar información más redundante en los criterios de evaluación, se realiza una selección de características basada en información mutua con coeficiente de correlación (Zhou, Wang, & Zhu , 2022).

Estas tres técnicas de la Tabla 12, fueron seleccionadas ya que describen ser las óptimas para representación de patrones en grandes conjuntos de datos. Nuestro enfoque se base en la utilización de técnicas de agrupación o clúster ing, debido a que esta técnica produjo informes que fueron simples de analizar y comprender durante las pruebas realizadas con el conjunto de datos académicos.

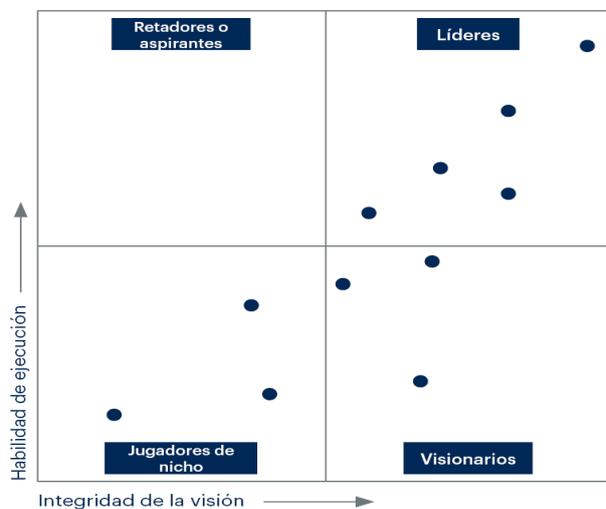
Al utilizar el algoritmo de agrupamiento no supervisado K-means, pudimos ver e identificar patrones en registros de modificación de notas que eran similares. Nuestro conjunto de datos se dividió en cuatro grupos utilizando la suma de cuadrados dentro del grupo como sugerencia para el número apropiado de clústeres (Njiru & Opiyo, 2018).

1.12.2. Cuadrante de Gartner

El Cuadrante Mágico de Gartner brinda una visión general amplia de las posiciones relativas de sus competidores y recopila los resultados de la investigación en un mercado en particular como se observa en la Figura 10, no se puede utilizar de acuerdo con el alcance declarado de tomar las técnicas de procesamiento de datos más populares de los últimos años debido a que no enumera técnicas avanzadas de análisis de datos según su clasificación únicamente clasifica plataformas y proveedores de análisis de datos.

Figura 10

Cuadrante Mágico de Gartner



Nota: Fuente (Gartner, 2023)

1.13. Metodología CRISP-DM

En el artículo (Pérez, Iturbide, & Olivares, 2015) señalan que debido a que los datos reales pueden ser erróneos, ruidosos o inconsistentes, la preparación de los datos

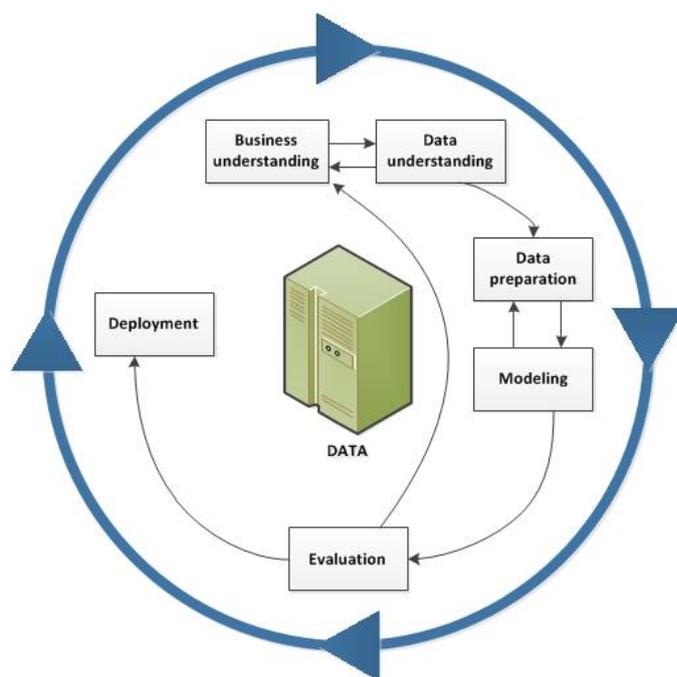
generalmente toma entre el 50% y el 70% del tiempo, pero ocasionalmente puede demorar hasta el 80 o el 90%.

Por lo que es importante la utilización de una metodología para procesar toda la información en grandes conjuntos de datos y descubrir el conocimiento en los datos, el CRISP-DM es una metodología de minería de datos que proviene de las siglas (Cross Industry Standard Process for Data Mining) (Núñez Arias, González Nuñez, Nonell Fernández, & Rodríguez Pupo, 2021).

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases como se observa en la Figura 11. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario (IBM, 2021).

Figura 11

Metodología CRISP-DM



Nota: Fuente (IBM, 2021)

La metodología CRISP-DM sigue siendo la norma de facto para la creación de proyectos de minería de datos y descubrimiento de conocimiento, según numerosas

encuestas y sondeos de usuarios (Plumed, Contreras Ochando, Ferri, & Hernández Orallo, 2021). en la Tabla 14 se detalla las fases de la metodología CRISP-DM.

Tabla 14

Fases de metodología CRISP-DM

Comprensión del negocio	Esta primera etapa se enfoca en comprender los objetivos y especificaciones del proyecto desde un punto de vista comercial y usar esa comprensión para definir el problema de minería de datos y crear un plan aproximado para alcanzar los objetivos.
Comprensión de los datos	La fase de comprensión de datos comienza con la recopilación inicial de datos y continúa con actividades para familiarizarse con los datos, identificar problemas de calidad de los datos, obtener información preliminar sobre los datos o identificar subconjuntos intrigantes para desarrollar hipótesis sobre información oculta.
Preparación de los datos	Las actividades para crear el conjunto de datos final a partir de los datos sin procesar iniciales se incluyen todas en la fase de preparación de datos. Entre las tareas se encuentran la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado.
Modelamiento	Durante esta fase se eligen y utilizan varias técnicas de modelado, y sus parámetros se calibran a valores ideales. Para el mismo tipo de problema de minería de datos, normalmente hay varias técnicas disponibles. Como resultado, con frecuencia se requiere volver a la etapa de preparación de datos.
Evaluación	En esta etapa del proyecto, ha creado un modelo (o modelos) que parecen ser de alta calidad desde la perspectiva del análisis de datos. Para asegurarse de que el modelo satisfaga con éxito los objetivos comerciales, es crucial realizar evaluaciones adicionales

del modelo y revisar los procesos utilizados para construir el modelo antes de avanzar con su implementación final.

Despliegue

En la mayoría de los casos, el proyecto no termina con la creación del modelo. La etapa de implementación puede variar en complejidad desde implementar un proceso de minería de datos repetible hasta algo tan simple como generar un informe, según los requisitos.

Nota: Fuente (Leventhal, 2010)

CAPÍTULO II

Desarrollo

Para el desarrollo del segundo capítulo se requirió utilizar un mayor volumen de datos del que se tenía previsto en el segundo objetivo específico, debido a que la cantidad de datos de la carrera de software de la UTN no brindaban resultados con mayor relevancia en la aplicación de técnicas avanzadas de procesamiento y análisis de datos por lo que se utilizó datos de las facultades de FACAÉ, FECYT, FICAYA, FICA y FCCSS del periodo académico SEP2022-FEB2023.

2.1. Aplicación del Metodología CRISP- DM

Los datos recabados del reporte de modificación de notas de las carreras de la UTN son utilizados en la ejecución de este proyecto utilizando la metodología CRISP-DM, en donde se realizan una limpieza de datos y aplicación de técnicas de minería de datos.

2.1.1. *Comprensión del negocio*

Sistema Académico UTN.

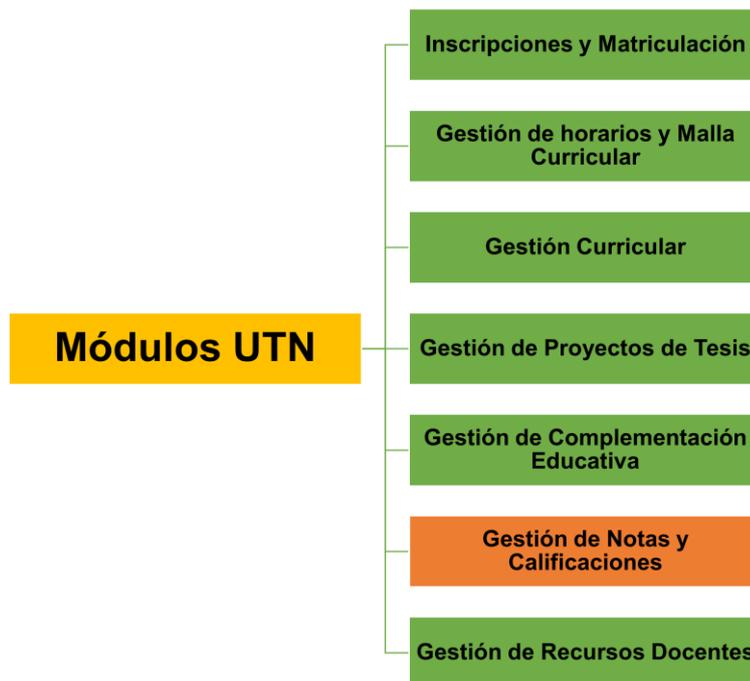
El sistema académico es una herramienta que facilita la gestión y estructuración de datos e información relacionada con el personal docente, administrativo y estudiantil de la institución.

La Universidad Técnica de Norte no es ajena a este hecho por esto se ha visto en la necesidad de trasladar toda la información y sistemas de la institución, a otra plataforma reduciendo así los costos e incrementando la capacidad de almacenamiento en Oracle Cloud para que los usuarios accedan a los recursos que están en línea a través de un servidor remoto (Hernández Obando, 2019).

Los módulos que conforman el sistema académico UTN se muestran en la Figura 12:

Figura 12

Módulos UTN



Misión.

La dirección de desarrollo tecnológico e informático de la Universidad Técnica del Norte, le corresponde administrar los servicios de informática, computación y comunicaciones, sin perjuicio de las demás funciones que se le recomiende. Ser el ente regulador de las políticas y normativas de carácter institucional; que deben ser llevadas a cabo con rigor, manteniendo el alto espíritu de calidad en todos los funcionarios, con el fin de lograr las expectativas encomendadas al departamento (UTN, 2013) (Hernández Obando, 2019).

Visión.

Establecer el rumbo estratégico del departamento y ejercer el liderazgo a nivel institucional, regional y nacional en el campo de la informática, computación y telecomunicaciones con tecnología de punta, investigaciones de avanzada e innovación que aportara para la transformación de la UTN (UTN, 2013) (Hernández Obando, 2019).

2.1.2. Comprensión de los Datos

Captura de Datos.

Para la obtención de los datos se realizó el proceso de solicitud de reporte de modificación de notas en el departamento de tecnologías de la información de la Universidad Técnica del Norte, los cuales fueron entregados en cinco documentos Excel separados por facultades.

Describir los Datos.

Los datos empleados para el estudio de este proyecto provienen del reporte de modificación de notas del periodo SEP2022-FEB2023 del módulo de gestión de notas y calificaciones de las facultades FACAE, FCCSS, FECYT, FICA, FICAYA de la Universidad Técnica del Nore. Los cuales se adjuntaron en un solo documento Excel formado 11 columnas y 58624 filas con datos tipo, entero, decimal, cadena de caracteres y fecha, los que se pueden observar en la Tabla 15.

Tabla 15

Estructura de reporte de modificación de notas

Columnas	Tipo de Dato
PERIODO	Cade de caracteres
NIVEL_CODIGO	Entero
FACULTAD	Cade de caracteres
CARRERA	Cade de caracteres
ASIGNATURA	Cade de caracteres
NOTA1	Decimal
NOTA2	Decimal
NOTA_SUPLETORIO	Decimal
NOTA_FINAL	Decimal
DOCENTE	Cade de caracteres
FECHAS_REGISTRO	Fecha, Cadena de caracteres, Decimal

Exploración de los Datos.

Mediante una exploración visual se puede observar que los datos indican en cada fila el registro y la modificación de notas de cada estudiante, organizado por Periodo, Nivel, Facultad, Carrera, Asignatura, Nota1, Nota2, Nota Supletorio, Nota Final, Docente y Fecha de Registro.

2.1.3. Preparación de los datos

Selección de los Datos.

Seleccionamos y nos centramos en la columna FECHA_REGISTRO que es la de mayor interés, ya que esta indica la fecha de registro seguida de la etiqueta del número de notas modifica (N1, N2, N3) y la nota modificada. La etiqueta "N1" hace referencia a la columna NOTA1, la etiqueta "N2" se refiere a la columna NOTA2 y la etiqueta "N3" se refiere a la columna NOTA_SUPLETORIO en caso de que tenga una nota supletorio.

Limpieza de Datos.

La herramienta Excel se utilizó para la limpieza de datos, el primer paso fue eliminar cadena de caracteres que no suman valor relevante en la columna FECHA_REGISTRO, de igual manera se eliminaron filas que contengan valores únicos tipo fecha como en la columna FECHA_REGISTRO.

Construcción de los Datos.

A partir de la columna FECHA_REGISTRO se delimito en columnas, para mantener los datos ordenados, formando un total de 62 registros.

Los registros fueron separados por Fecha, Etiqueta (N1, N2, N3) y Nota, para mantener los datos ajustados al momento de aplicar la técnica de minería de datos.

Integración de Datos.

Los datos que se generaron a partir de la columna FECHA_REGISTRO como se indica en la Tabla 18, se integran con el resto de las columnas Periodo, Nivel, Facultad, Carrera, Asignatura, Nota1, Nota2, Nota Supletorio, Nota Final, Docente, como se muestra

en la Tabla 12 y Tabla 13 para una mejor comprensión de los datos formando un total de 196 columnas y 56511 filas.

Formato de los Datos.

Por último, en se realiza una modificación sintáctica en las notas cambiando la coma decimal por un punto, se ordenó los datos por fecha y cambió el formato de estas utilizando Python de acuerdo con la implementación de la técnica y algoritmo a ser utilizado.

2.1.4. Modelamiento

Selección de la Técnica.

La agrupación en clústeres es una de las tareas más básicas e importantes en el análisis de datos, que tiene amplias aplicaciones. Se ha estudiado en varios campos de estudio, incluida la minería de datos, el aprendizaje automático, el reconocimiento de patrones y el análisis de datos biomédicos, económicos y de ingeniería (Simić, Banković, Villar, & Simić, 2020).

De acuerdo con el estudio realizado, la técnica que mejor se adapta al conjunto de datos y los objetivos de la investigación es clúster ing, técnica que será aplicada para agrupar datos similares utilizando el algoritmo KMeans con la librería sklearn en Python, para poder identificar patrones que ayuden encontrar evidencias o datos anómalos en el proceso de auditoría informática.

Diseño de las Pruebas.

El índice de Calinski-Harabasz se usa para estimar el número óptimo y correcto de conglomerados, mediante el uso de dos medidas conocidas como: el criterio de la razón de varianza y el total (Simić, Banković, Villar, & Simić, 2020). La librería sklearn también cuenta con esta función que puede ser importada para una fácil utilización en Python.

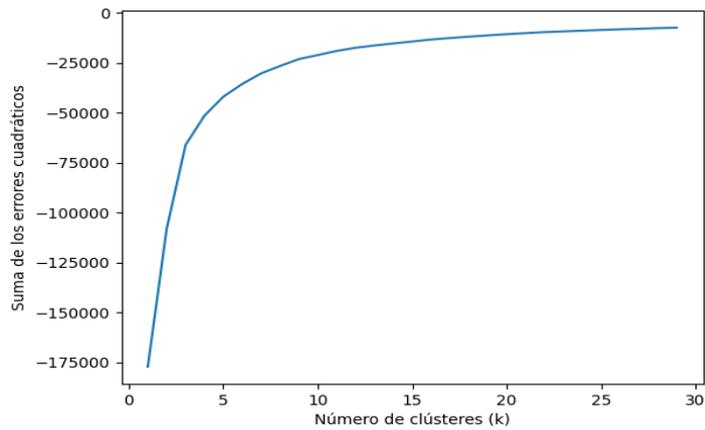
Construcción del Modelo.

En esta fase se pone en ejecución la técnica seleccionada de agrupamiento realizado con Python en la plataforma Google Colab formando diferentes clúster s con todas las

variables notas del conjunto de datos. Para determinar el valor óptimo de k , es decir el número de clúster, se utiliza la curva elbow como se muestra en la Figura 13, en la cual nos sugiere realizar un clúster con 4 agrupaciones.

Figura 13

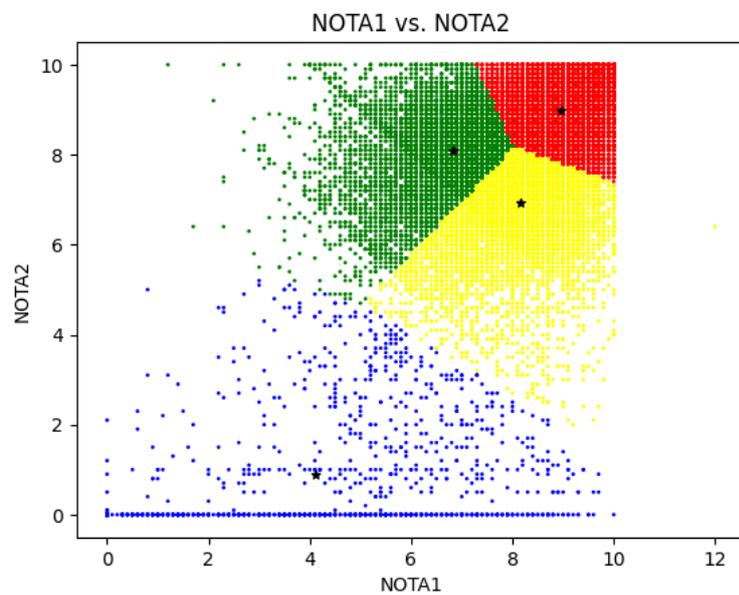
Curva elbow



Luego de estimar el número adecuado de clústeres se procede a la ejecución del algoritmo K-Means formando clústeres con dos variables cuantitativas de notas. El primer clúster de la Figura 14 sirvió como referencia para el análisis visual debido que está formado con las notas finales Nota1 y Nota2, los demás clústeres se formaron con los registros de notas modificados que se observa en las Figuras 14 - 24.

Figura 14

Clúster 1 Nota1 vs Nota2

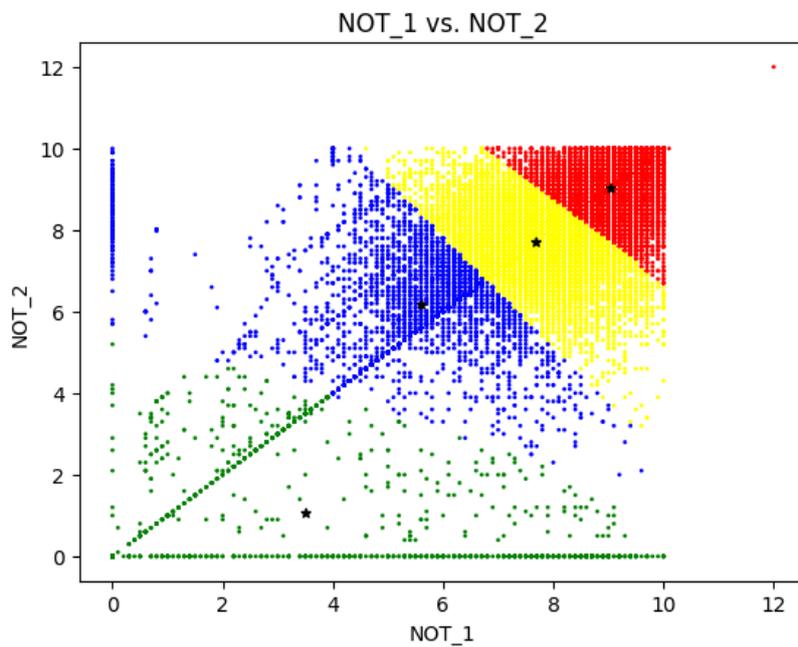


Nota: La Nota1 y la Nota2 representan a las notas finales del registro de notas

En la Figura 14 se observa la segmentación de clústers en la que la agrupación de color azul representa las notas bajo rendimiento, en la que se mira una dispersión muy amplia al resto de agrupaciones, la agrupación de color amarillo y azul representan las notas de rendimiento intermedio y la agrupación de color rojo representan las notas alto rendimiento.

Figura 15

Clúster 2 Not_1 vs Not_2

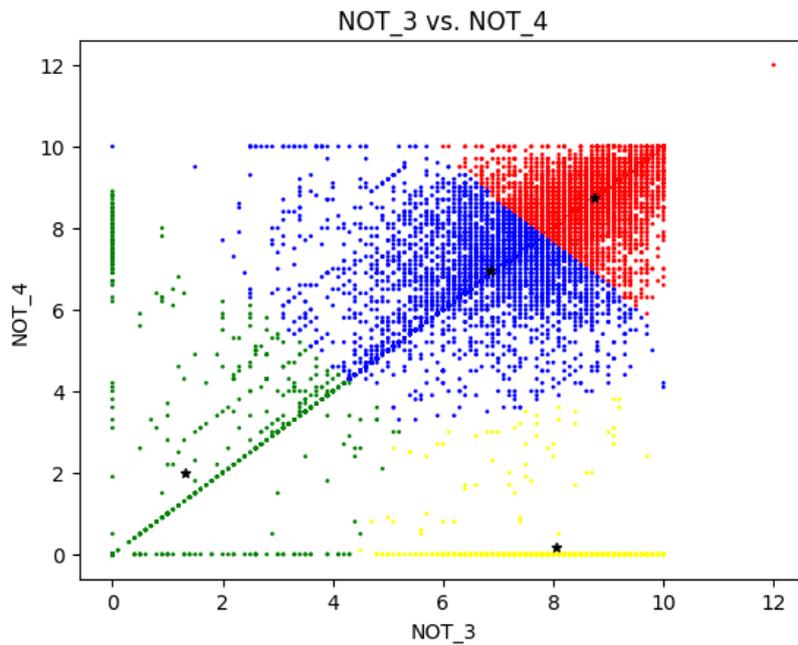


Nota: *Not_1* y *Not_2* representan a las notas del primer y segundo registró de modificación de notas

En la Figura 15 se observa la segmentación de clústers en la que la agrupación de color verde representa las notas bajo rendimiento, en la que se mira una dispersión bastante similar de la Figura 14, la agrupación de color azul representa las notas de rendimiento intermedio bajo, la agrupación de color amarillo representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento.

Figura 16

Clúster 3 Not_3 vs Not_4

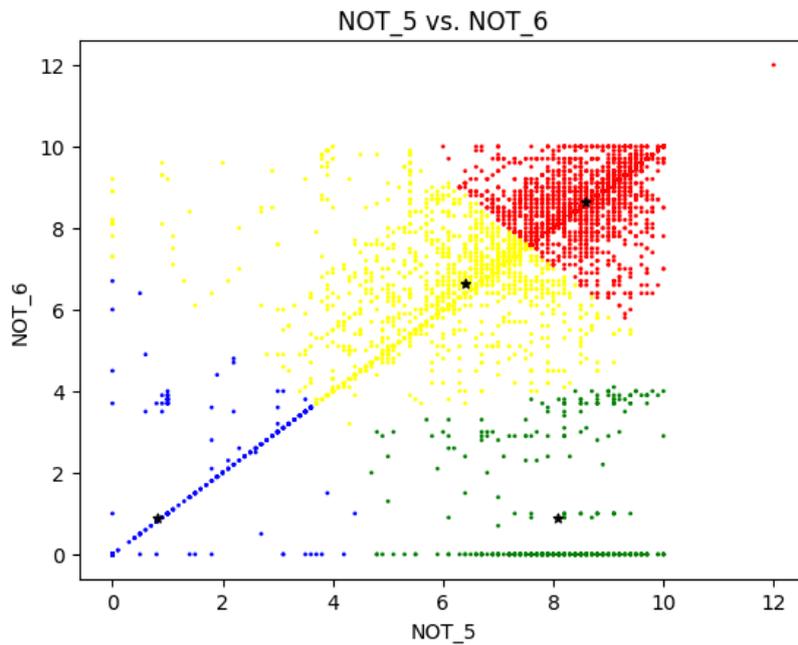


Nota: *Not_3 y Not_4 representan a las notas del tercer y cuarto registró de modificación de notas.*

En la Figura 16 se observa la segmentación de clústers en la que la agrupación de color verde representa las notas bajo rendimiento, en la que se mira una dispersión reducida en relación a las Figuras 14,15, la agrupación de color amarillo representa las notas de bajo rendimiento en relación con la nota_4, la agrupación de color azul representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento en la que se puede observar que el rango de dispersión aumentó levemente.

Figura 17

Clúster 4 Not_5 vs Not_6

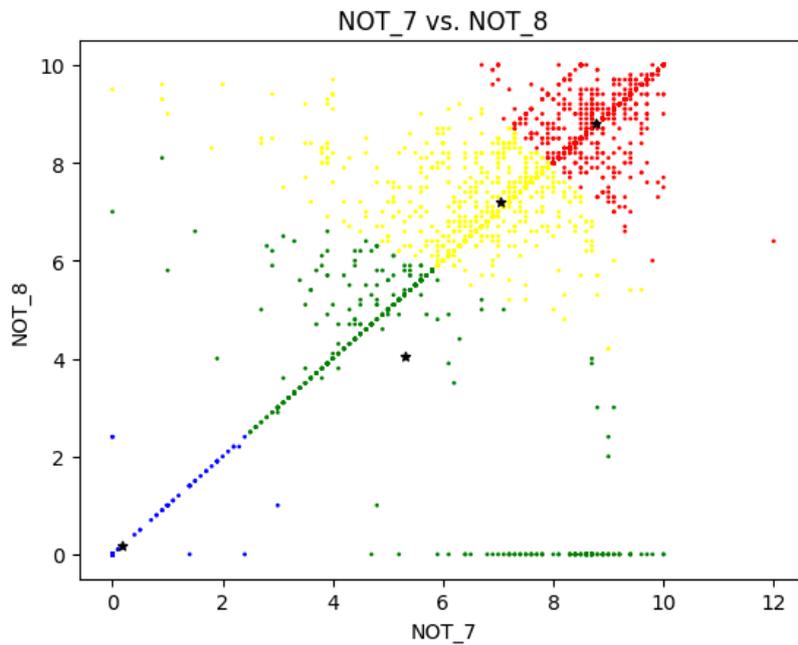


Nota: Not_5 y Not_6 representan a las notas del quinto y sexto registró de modificación de notas.

En la Figura 17 se observa la segmentación de clústers en la que la agrupación de color azul representa las notas bajo rendimiento, en la que se mira una dispersión reducida en relación con las Figuras 14,15,16, la agrupación de color verde representa las notas de bajo rendimiento en relación con la nota_6, la agrupación de color amarillo representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento similar a la Figura 16.

Figura 18

Clúster 5 Not_7 vs Not_8

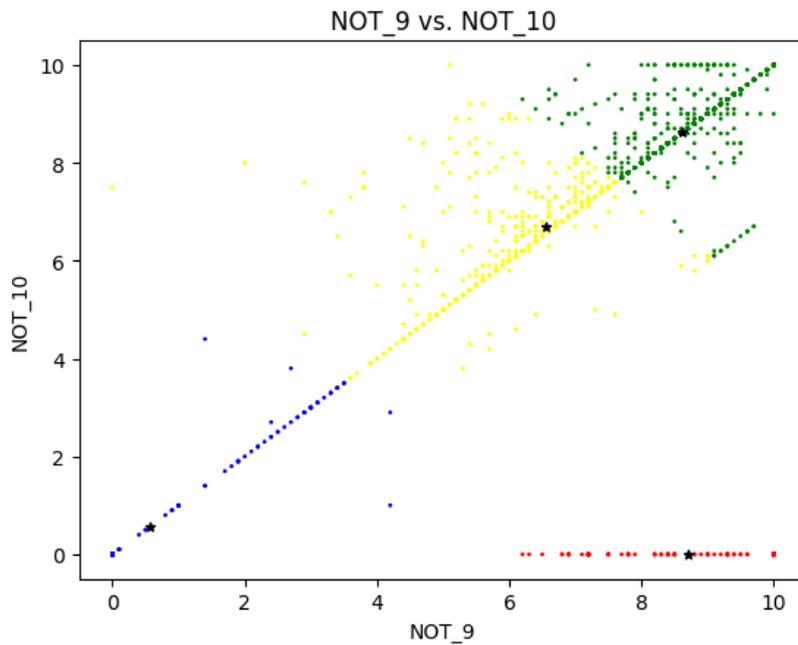


Nota: Not_7 y Not_8 representan a las notas del séptimo y octavo registró de modificación de notas.

En la Figura 18 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color azul representa las notas de bajo rendimiento, en la que se mira una dispersión casi nula, la agrupación de color verde representa las notas de rendimiento intermedio bajo, la agrupación de color amarillo representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento.

Figura 19

Clúster 6 Not_9 vs Not_10

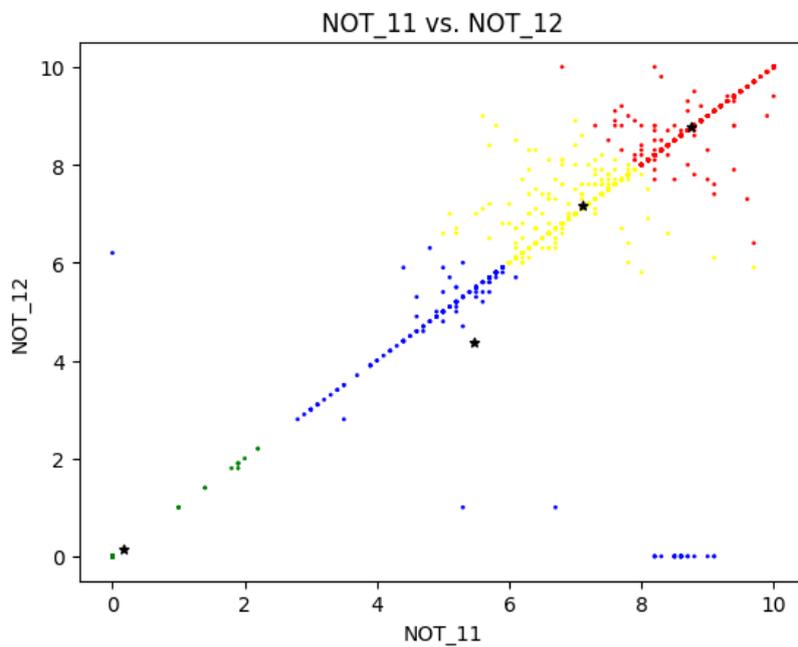


Nota: Not_9 y Not_10 representan a las notas del noveno y décimo registró de modificación de notas.

En la Figura 19 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color azul representa las notas bajo rendimiento, en la que se mira una dispersión bastante similar de la Figura 18, la agrupación de color rojo representa las notas de rendimiento intermedio bajo casi nula, la agrupación de color amarillo representa las notas de rendimiento intermedio alto y la agrupación de color verde representan las notas alto rendimiento.

Figura 20

Clúster 7 Not_11 vs Not_12

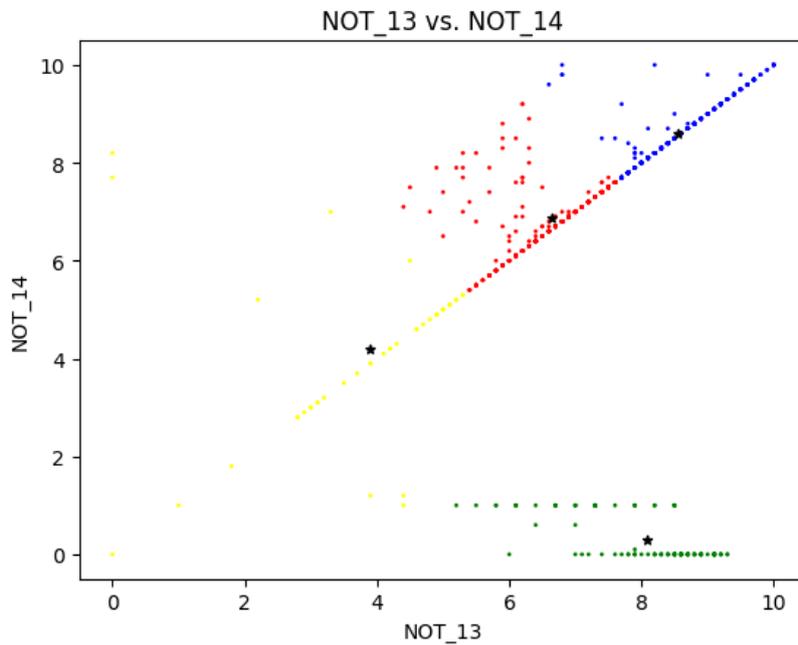


Nota: Not_11 y Not_12 representan a las notas del décimo primero y décimo segundo registró de modificación de notas.

En la Figura 20 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color verde representa las notas de bajo rendimiento, en la que se mira una dispersión bastante similar de la Figura 18,19, la agrupación de color azul representa las notas de rendimiento intermedio bajo, la agrupación de color amarillo representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento.

Figura 21

Clúster 8 Not_13 vs Not_14

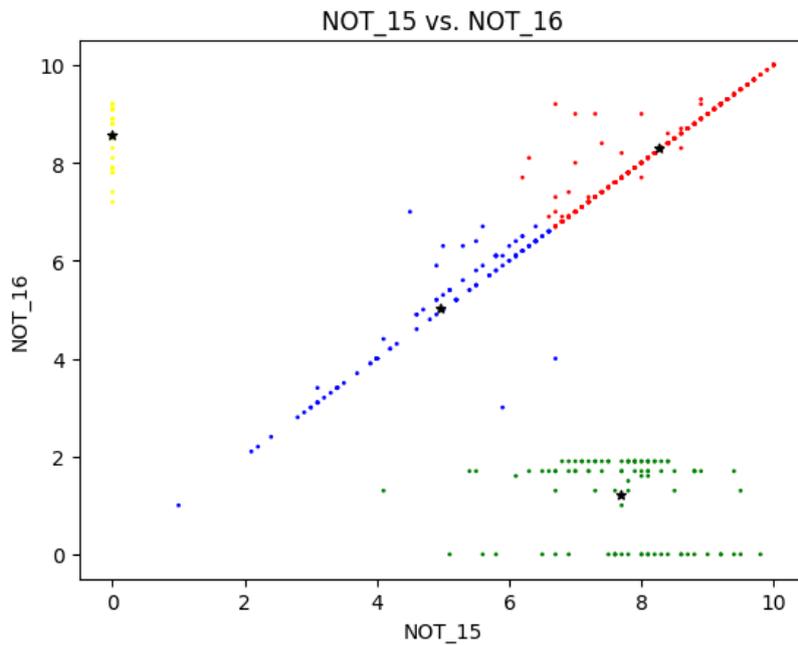


Nota: Not_13 y Not_14 representan a las notas del décimo tercero y décimo cuarto registró de modificación de notas.

En la Figura 21 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color amarillo representa las notas de bajo rendimiento, en la que se mira una el aumentó de dispersión de puntos, la agrupación de color verde representa las notas de rendimiento intermedio bajo casi nula en relación a la nota_14, la agrupación de color rojo representa las notas de rendimiento intermedio alto y la agrupación de color azul representan las notas alto rendimiento.

Figura 22

Clúster 9 Not_15 vs Not_16

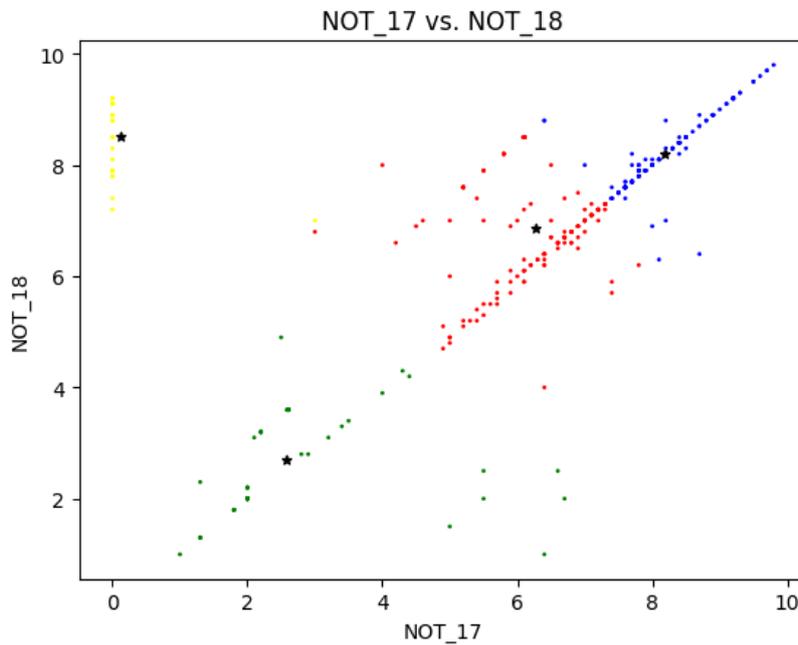


Nota: Not_15 y Not_16 representan a las notas del décimo quinto y décimo sexto registro de modificación de notas.

En la Figura 22 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color azul representa las notas de rendimiento intermedio, la agrupación de color amarillo representa las notas de bajo rendimiento con relación a la nota_15, la agrupación de color verde representa las notas de bajo rendimiento en relación con la nota_16 y la agrupación de color rojo representan las notas alto rendimiento.

Figura 23

Clúster 10 Not_17 vs Not_18

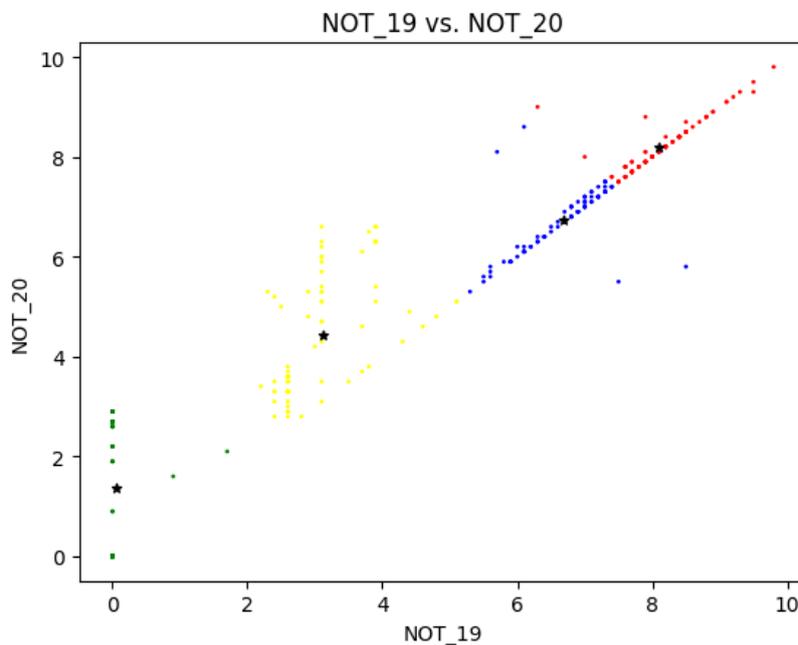


Nota: Not_17 y Not_18 representan a las notas del décimo quinto y décimo sexto registro de modificación de notas.

En la Figura 23 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color verde representa las notas de bajo rendimiento, la agrupación de color rojo representa las notas de rendimiento intermedio, la agrupación de color amarillo representa las notas de bajo rendimiento en relaciona la nota_17 y la agrupación de color azul representan las notas alto rendimiento.

Figura 24

Clúster 11 Not_19 vs Not_20



Nota: Not_19 y Not_20 representan a las notas del décimo noveno y veinteavo registro de modificación de notas.

En la Figura 24 se observa la segmentación de clústers con la dispersión de puntos reducida en la que la agrupación de color verde representa las notas bajo rendimiento, en la que se mira una dispersión bastante similar de la Figura 18,19,20, la agrupación de color amarillo representa las notas de rendimiento intermedio bajo, la agrupación de color azul representa las notas de rendimiento intermedio alto y la agrupación de color rojo representan las notas alto rendimiento.

Se realiza el agrupamiento de datos hasta el registro 20 de 62 registros, debido que el número de datos de los siguientes registros disminuyen considerablemente y no brinda un valor óptimo en la evaluación de conglomerados.

Dado que el índice de Calinski-Harabasz es una relación entre la suma de cuadrados dentro del denominador y el numerador, las soluciones de conglomerados con valores más altos del índice son "mejores" que las soluciones de conglomerados con valores más bajos (Alteryx, 2019).

2.1.5. Evaluación

El método Calinski-Harabasz es una forma de evaluar el agrupamiento interno del clúster, entre más alto sea el valor, mejor es el agrupamiento. En la Tabla 22 se indica el valor de cada clúster ing al aplicar el método Calinski-Harabasz.

Tabla 16

Evaluación de clúster ing

Clústeres	Valor Calinski-Harabasz
Clúster 1	69959.511
Clúster 2	78849.498
Clúster 3	45627.944
Clúster 4	27557.290
Clúster 5	17350.206
Clúster 6	11582.910
Clúster 7	6539.960
Clúster 8	2666.372
Clúster 9	1303.220
Clúster 10	1226.835
Clúster 11	1467.924

En la Tabla 22 se observa que el nivel de agrupamiento de cada clúster disminuye al aplicar a las variables con datos reducidos, tal como se visualiza en las Figuras 14 - 24 de dispersión de puntos de clúster.

2.1.6. Despliegue

La fase de despliegue se realiza un reporte de resultados a la institución, aplicada al capo de auditoría informática.

2.2. Planificación de Auditoría (Fase I).

Se tiene como finalidad realizar una auditoría informática para verificar la situación actual de la calidad de los datos del sistema académico de la universidad técnica del norte en el módulo de gestión de notas y calificaciones utilizando una técnica avanzada de análisis de datos, basándose en la norma ISO/IEC 25012:2008.

2.2.1. Contacto con el cliente.

El sujeto de la auditoría es la Universidad Técnica del Norte, Departamento de Informática, Sistema Integrado informático Universitario- sistema académico.

2.2.2. Estudio Inicial del Entorno a Auditar.

Este plan de auditoría se desarrollará con los reportes de datos históricos de cambios en las notas relacionadas con las carreras que conforman la Universidad Técnica del Norte, además se realizó una encuesta y entrevista al DDTI para conocer la situación actual de la calidad de los datos.

2.2.3. Equipo Auditor

- Imbaquingo Esparza Daisy Elizabeth (Auditor Líder)
- Alvarez Osorio Francisco Javier (Auditor Junior)

2.2.4. Plan Preliminar

Antecedentes.

La UTN como referente de la región norte del Ecuador desarrolla sus aplicaciones informáticas y las integra al SIIU, la cantidad de información aumenta exponencialmente con el tiempo, la seguridad de los sistemas informáticos también han aumentado siendo expuesta a ataques y en efecto a la vulnerabilidad de los datos.

Objetivo general

Realizar una auditoría informática de los datos del sistema académico de la UTN aplicando técnicas avanzadas de análisis de datos para determinar su calidad.

Objetivos Específicos

- Recolección y limpieza de datos del módulo de gestión de notas y calificaciones.
- Analizar los datos con técnicas de minería de datos.
- Elaborar un informe técnico con recomendaciones en base a los resultados obtenidos con la norma ISO/IEC 25012:2008

Alcance

Para realizar la auditoría del sistema académico de la Universidad del Norte en el módulo de gestión de notas y calificaciones, se toma en cuenta la norma ISO/IEC 25012:2008 para comprobar la calidad de los datos inherente y dependiente del sistema como la confidencialidad, precisión, trazabilidad y comprensibilidad, aplicando técnica avanzada de análisis de datos para descubrimiento de patrones en el registro de modificación de notas, de las facultades FACAE, FCCSS, FECYT, FICA, FICAYA, del periodo SEP2022-FEB2023.

Recursos para la auditoría

Tabla 17

Recursos de auditoría

RECURSOS
Recursos humanos
▪ Auditor, Equipo Auditor y Cliente
Equipos tecnológicos
▪ Computador
Otros
▪ Python, Minería de Datos, Google Colab

Cronograma

Cronograma																			
Se ha programado que la auditoría se la realizará en un tiempo estimado de 12 semanas, las cuales se contará desde el día 10 de abril de 2023.																			
ACTIVIDADES	RESULTADOS POR OBJETIVO / ACTIVIDAD	CRONOGRAMA DE ACTIVIDADES															% de avance / actividad	% de cumplimiento / objetivo	% avance TOTAL
		SEMANAS 2023																	
		15	16	17	18	19	20	21	22	23	24	25	26						
Contacto con el cliente	brindar información	■															100%	100%	100%
Estudio Inicial del entorno a auditar	Conocer la institución	■															100%	100%	100%
Definir el equipo auditor	Asesoramiento de auditoría	■															100%	100%	100%
Elaboración del Plan Preliminar	Informe preliminar	■	■	■													100%	100%	100%
Elaboración de propuesta de auditoría	Recomendaciones a la institución		■	■													100%	100%	100%
Contrato	Establecer cumplimientos		■	■													100%	100%	100%
Ejecución de la auditoría	Verificar situación actual			■	■	■	■	■	■	■							100%	100%	100%
Validación de la auditoría	Evaluación de auditoría									■	■	■					100%	100%	100%
Seguimiento	Informe final											■	■	■			100%	100%	100%

Costos

En la Tabla 18 se muestra la descripción de costos de herramientas y servicios utilizados en la auditoría.

Tabla 18

Descripción de Costos

DESCRIPCIÓN	COSTO ESTIMADO
HARDWARE	
Computadora	\$ 800.00
SOFTWARE	
Sistema Operativo	\$ 145.00
SERVICIOS BÁSICOS	
Internet	\$ 25.00
Luz	\$ 14.00
MATERIALES DE OFICINA	
Papel Bond	\$ 3.00
INVESTIGACION	
Impresión de Solicitudes	\$ 1.00
SUBTOTAL	\$ 988
10% DE IMPREVISTOS	\$ 98.8
TOTAL	\$ 1.086.8

Riesgos de la auditoría

En la Tabla 19 se muestra los riesgos en la ejecución de la auditoría informática como son inherentes, de control y de detección.

Tabla 19

Riesgos de auditoría

RIESGOS		
INHERENTES	DE CONTROL	DE DETECCIÓN
1. Factores externos.		1. Habilidades blandas del equipo auditor.
2. Objetivos y estrategias.		2. Objetivos y alcance claros.
3. Control interno.		3. Aptitud y actualidad del equipo auditor.
4. Disponibilidad de tiempo en los encargados de la IES.	1. Supervisión. 2. Comunicación. 3. Mecanismo de control de gestión u operación.	4. Enfermedad o percances en equipo auditor.
5. Falta de documentación.	4. Ambiente o entorno de control.	5. Conocimiento previo de la IES para auditar.
6. Causas de incorrecciones.	5. Evaluación de riesgos.	6. Conocimiento de técnicas de auditoría.
7. Información fraudulenta		7. Compresión del MAIIES.
8. Tamaño de la IES		

2.2.5. Propuesta de Auditoría

Descripción del problema. El crecimiento de la información en los sistemas académicos de las instituciones de educación superior dificulta el análisis de grandes volúmenes de datos durante las auditorías informáticas, lo que genera ineficiencia y resultados subjetivos. Como estudio de caso práctico, se evaluará el SIIU para valorar la calidad de los datos del módulo de gestión de notas.

Solución Propuesta. Debido al problema propuesto se realizará una auditoría informática para identificar patrones que sean relevantes en un gran volumen de datos, aplicando técnicas avanzadas de análisis de datos, con el objetivo de mejorar la toma de decisiones durante la auditoría y emitir un informe técnico objetivo y eficiente al finalizar la misma, proponiendo soluciones que ayude a mejorar la calidad de los datos como la precisión, trazabilidad y confidencialidad que se especifican en la norma ISO/IEC 25012:2008.

Ventajas solución de la propuesta. El desarrollo de esta auditoría informática permite evaluar todo el conjunto de datos puesto a consideración, para visualizar patrones e identificar posibles fallas y vulnerabilidades a las que se han visto expuestos los sistemas académicos, y el impacto que podrían causar en las instituciones. Conjuntamente, el departamento informático recibiera un informe técnico con recomendaciones para la correcta toma de decisiones.

2.2.6. Contrato.

El contrato de auditoría normalmente tiene diez cláusulas, el cliente y el auditor son responsables de reducir o aumentar el número de cláusulas. Debido a que esta auditoría es parte de un proyecto para titulación la aprobación del tema se toma en cuenta como un tipo de contrato en esta auditoría.

2.3. Ejecución de la Auditoría (Fase II)

En la Tabla 20 se muestra el plan de trabajo dividido por actividades, instrumentos de investigación, fuente, fecha de aplicación, responsable, observaciones y resultados.

2.3.1. Plan de Trabajo

Tabla 20

Plan de trabajo

MÉTODO DE AUDITORÍA INFORMÁTICA PARA INSTITUCIONES DE EDUCACIÓN SUPERIOR (MAIIES)							
PLAN DE TRABAJO							
N°	Actividad	Instrumento de Investigación	Fuente	Fecha de aplicación	Responsable	Observaciones	Resultados
1	Contacto con el cliente	Resolución HCD Nro. UTN-FICA-2021-08-010	Consejo Directivo de la Facultad de Ingeniería en Ciencias Aplicadas	10/04/2023	MSc. Imbaquingo Esparza Daisy Elizabeth	Ninguna	Aprobación de auditoría módulo de gestión de notas y calificaciones

2	Estudio del entorno a auditar	Entrevista	Elaboración propia	11/04/2023	Sr. Alvarez Osorio Francisco Javier	Ninguna	Conocer situación actual de la calidad de los datos
3	Definir equipo de auditoría	Acta de reunión 001	Elaboración propia	12/04/2023	Sr. Alvarez Osorio Francisco Javier MSc. Imbaquingo Esparza Daisy Elizabeth	Ninguna	Realización, seguimiento y asesoramiento de la auditoría
4	Elaboración del Plan Preliminar	Documento plan preliminar	Método de Auditoría Informática para Instituciones de Educación	13/04/2023	Sr. Alvarez Osorio Francisco Javier	Ninguna	Informe preliminar

			Superior (MAIIES)				
			Método de Auditoría				
5	Elaboración de propuesta de auditoría	Documento de propuesta de auditoría	Informática para Instituciones de Educación Superior (MAIIES)	19/04/2023	Sr. Alvarez Osorio Francisco Javier	Ninguna	Recomendaciones a la institución
			Método de Auditoría		Sr. Alvarez Osorio Francisco Javier		
6	Contrato	Contrato de auditoría	Informática para Instituciones de Educación Superior (MAIIES)	21/04/2023	MSc. Imbaquingo Esparza Daisy Elizabeth	Ninguna	Establecer cumplimientos

2.3.2. Investigación de Campo

Durante la investigación de campo se recopiló información aplicando una encuesta y entrevista al DDTI también se solicitó los datos del informe de modificación de notas del periodo SEP2022-FEB2023 para utilizar técnicas avanzadas de análisis de datos e identificar patrones

Encuesta

Para la investigación de campo se aplicó una encuesta al DDTI a los encargados de desarrollo y seguridad, se utilizaron las siguientes preguntas que se muestran en la Tabla 20.

Tabla 21

Preguntas de encuesta

N°	Preguntas
1	¿La Institución de Educación Superior IES tiene políticas para la seguridad de la información?
2	¿La Institución de Educación Superior IES cuenta con un responsable de Seguridad de la información?
3	¿Las políticas para la seguridad de la información están aprobadas por la dirección, publicadas y comunicadas a los empleados?
4	¿Las políticas para la seguridad de la información son revisadas con regularidad, y tienen mejoras significativas para garantizar su idoneidad, adecuación y efectividad?
5	¿Existen políticas y un proceso formal disciplinario de seguridad de la información comunicado a empleados, y monitoreo de cumplimiento?
6	¿Se ha desarrollado e implementado un conjunto apropiado de procedimientos para el etiquetado y tratamiento de la información, de acuerdo

	con el esquema de clasificación adoptado por la Institución de Educación Superior IES?
7	¿Se han establecido, documentado, revisado y aprobado una política de control de accesos en base a las necesidades de seguridad y de negocio de la Institución de Educación Superior IES?
8	¿Existe un control de acceso a las redes y servicios asociados?
9	¿Existe un procedimiento formal de altas y bajas de usuarios con objeto de habilitar la asignación de derechos de acceso?
10	¿Se ha implementado un proceso formal de aprovisionamiento de accesos a los usuarios para asignar o revocar derechos de acceso a todos los tipos de usuarios y para todos los sistemas y servicios?
11	¿La asignación y uso de derechos de acceso con privilegios especiales son restringidos y controlados?
12	¿La asignación de información confidencial para la autenticación es supervisada mediante un proceso de gestión controlado?
13	¿Se exige a los usuarios el uso de las buenas prácticas de seguridad en el uso de información confidencial para la autenticación?
14	¿Se restringe el acceso a los usuarios y al personal de mantenimiento de sistemas y aplicaciones, en relación con la política de control de accesos definida?
15	¿Tienen procedimientos seguros de inicio de sesión?
16	¿Tienen un manejo adecuado de gestión de contraseñas de usuario?
17	¿Se dispone de herramientas de administración de sistemas y estas son restringidas y estrechamente controladas?
18	¿Existe una restricción del acceso al código fuente de las aplicaciones software?

19	¿Se protege contra posibles alteraciones y accesos no autorizados la información de los registros?
20	¿Se registran las actividades del administrador, del operador del sistema y los registros asociados son revisados de manera regular?
21	¿Se han planificado y acordado los requisitos y las actividades de auditoría que involucran la verificación de los sistemas operacionales?

Entrevista

El propósito de la entrevista con el director del departamento de tecnología de la información fue comprender los procesos académicos y la operación del sistema con respecto a la precisión, trazabilidad y comprensibilidad de la calidad de datos establecidos en la norma ISO/IEC 25012: 2008, preguntas que se detallan en la Tabla 21.

Tabla 22

Preguntas de entrevista

N°	Pregunta
1	¿En la institución que usted se encuentra laborando actualmente qué tipo de auditoría se realiza?
2	¿Si su respuesta a la pregunta anterior fue afirmativa, indique si la auditoría fue realizada por auditores internos o externos y qué áreas del departamento fueron auditadas?
3	¿Cada que tiempo y bajo qué motivos se realizan auditorías informáticas en el departamento de tecnologías de su institución?
4	¿Dentro del proceso de auditoría se utilizaron técnicas de avanzadas de análisis de datos para identificar patrones en los datos, Cuáles fueron las técnicas aplicadas?

5	¿Cómo se controla y se mide la confidencialidad de los datos del sistema académico de la institución?
6	¿Cómo se garantiza la integridad y la confiabilidad de los datos a lo largo de todo el ciclo de vida del sistema académico?
7	¿Puede describir quiénes tienen acceso y la autoridad para subir y modificar notas y calificaciones?
8	¿Cuál es el proceso que se debe llevar para modificar notas en el módulo de gestión de notas y calificaciones?
9	¿Cuál es el tiempo administrado por el sistema para subir y modificar notas?
10	¿Bajo qué motivos se permite el modificar notas después de culminada la fecha de asentamiento de notas?
11	¿Qué prácticas o técnicas implementa para asegurar la precisión de los datos del sistema académico?
12	¿Según su experiencia cuales cree que son los protocolos o procedimientos establecidos para resolver inconsistencias de imprecisión identificadas en el sistema académico?
13	¿Se lleva a cabo alguna auditoría o revisión periódica para verificar la trazabilidad del sistema académico?
14	¿Se realiza un seguimiento y registro de todas las modificaciones realizadas en el sistema académico? ¿Cómo se documentan estas modificaciones?
15	¿Existe algún mecanismo para rastrear la relación entre los datos ingresados en el sistema académico y los resultados generados a partir de ellos?

16 ¿Se ofrecen capacitaciones o sesiones de formación para los usuarios del sistema académico, con el fin de mejorar su comprensión y aprovechamiento de sus características?

17 ¿Se recopilan y analizan los comentarios y sugerencias de los usuarios para mejorar la comprensibilidad del sistema académico en futuras iteraciones o versiones?

Reporte de notas

Se solicitaron datos en DDTI del reporte de modificación de nota de las facultades FACA E, FCCSS, FECYT, FICA y FICAYA de la Universidad Técnica del Norte correspondiente al período SEP2022-FEB2023 por medio del Memorando Nro. UTN-FICA-CSOFT-0146-M.

2.3.3. Análisis de la información recolectada

En esta fase se tabula las encuestas y entrevistas realizadas, se hace un análisis de clústeres del reporte de modificación de notas y se procede a un análisis descriptivo de la Confidencialidad, Precisión, Trazabilidad, Comprensibilidad de los datos en base a la norma ISO/IEC 25012:2008.

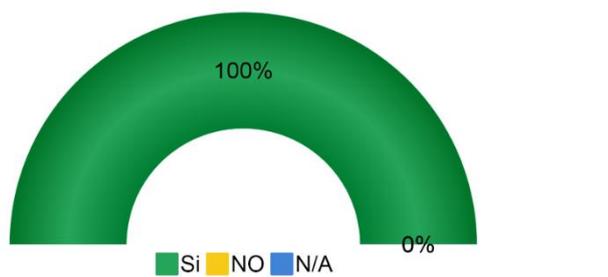
Tabulación de la encuesta

Para la aplicación de la auditoría informática, se recopiló información mediante una encuesta formulada al DDTI a los encargados de desarrollo y seguridad de la Universidad Técnica del Norte. Con el fin de verificar políticas de seguridad y verificar la situación actual de calidad de los datos. A continuación, se detalla la tabulación de cada una de las preguntas mostradas en la Tabla 20.

Figura 25

Pregunta 1

¿La Institución de Educación Superior IES tiene políticas para la seguridad de la información?

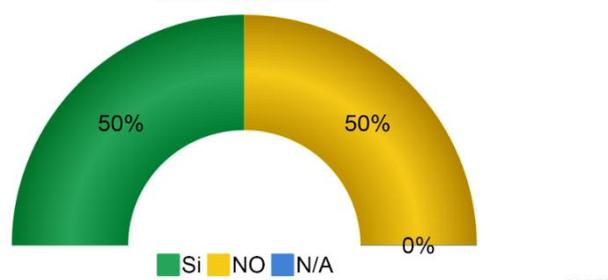


En la Figura 25 se muestra que los encuestados están de acuerdo en un 100% que la UTN cuenta con políticas para a seguridad de información.

Figura 26

Pregunta 2

¿La Institución de Educación Superior IES cuenta con un Responsable de Seguridad de la información?

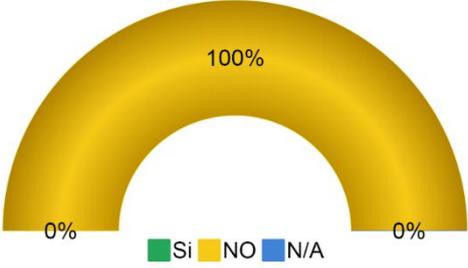


En la Figura 26 se muestra que el 50% de los encuestados no cuenta con responsable de la seguridad de la información mitras que el 50% opina que si cuenta con un responsable de la información.

Figura 27

Pregunta 3

¿Las políticas para la seguridad de la información están aprobadas por la dirección, publicadas y comunicadas a los empleados?

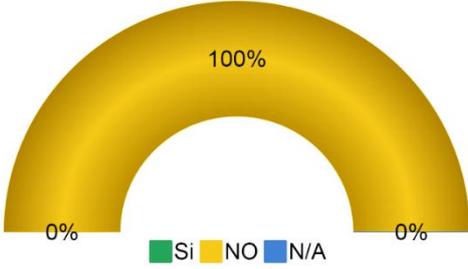


En la pregunta 27 se observa que el 100% de los encuestados afirma que las políticas para la seguridad de la información no están aprobadas por la dirección, publicadas y comunicadas a los empleados.

Figura 28

Pregunta 4

¿Las políticas para la seguridad de la información son revisadas con regularidad, y tienen mejoras significativas para garantizar su idoneidad, adecuación y efectividad?

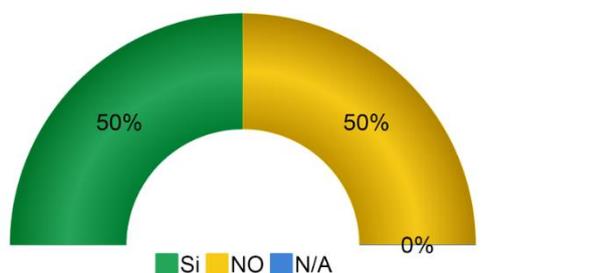


En la Figura 28 el 100% de los encuestados afirma que las políticas para la seguridad no son revisadas con regularidad.

Figura 29

Pregunta 5

¿Existen políticas y un proceso formal disciplinario de seguridad de la información comunicado a empleados, y monitoreo de cumplimiento?

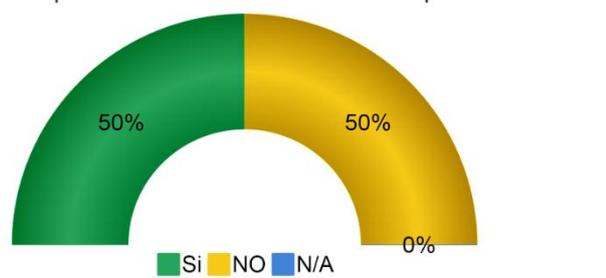


En la Figura 29 un 50% de los encuestados afirma que si cuenta con políticas y un proceso disciplinario de seguridad de la información y el otro 50% dice que no.

Figura 30

Pregunta 6

¿Se ha desarrollado e implementado un conjunto apropiado de procedimientos para el etiquetado y tratamiento de la información, de acuerdo con el esquema de clasificación adoptado por la Institución de Educación Superior IES?

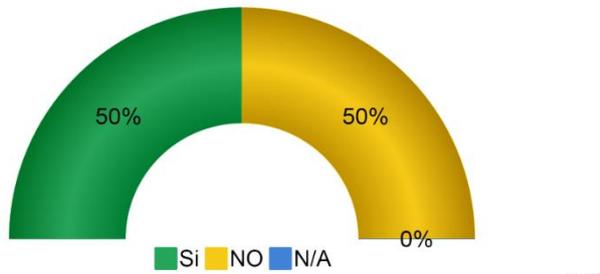


En la Figura 30 se observa que el 50% de los encuestados dice que si se ha desarrollado e implementado un conjunto apropiado de procedimientos para el etiquetado y tratamiento de la información y el otro 50% dice que no.

Figura 31

Pregunta 7

¿Se han establecido, documentado, revisado y aprobado una política de control de accesos en base a las necesidades de seguridad y de negocio de la Institución de Educación Superior IES?

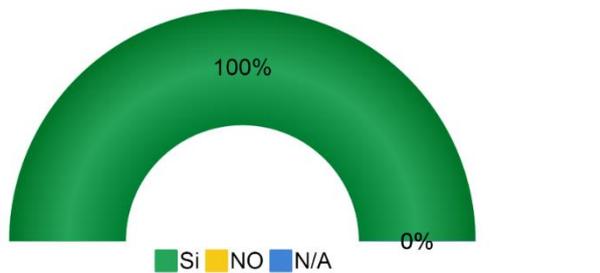


En la Figura 31 se observa que el 50% de los encuestados dice que si se ha establecido y aprobado políticas de control de acceso en base a las necesidades de seguridad y el otro 50% dice que no.

Figura 32

Pregunta 8

¿Existe un control de acceso a las redes y servicios asociados?

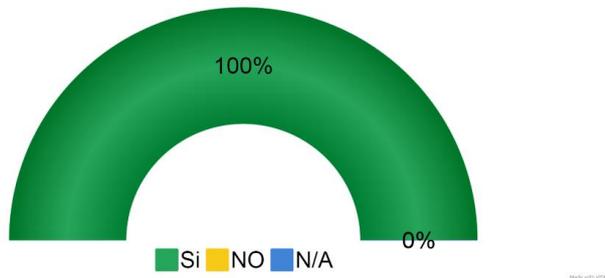


En la Figura 32 se observa que el 100% de los encuestados coincide que si existe un control de acceso a las redes y servicios asociados.

Figura 33

Pregunta 9

¿Existe un procedimiento formal de altas y bajas de usuarios con objeto de habilitar la asignación de derechos de acceso?

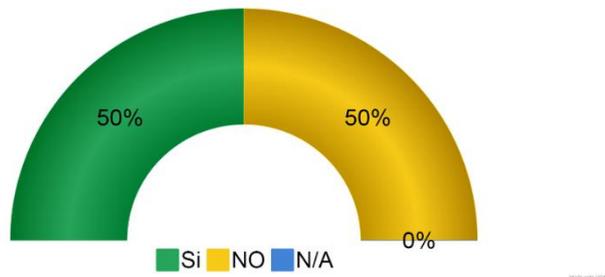


En la Figura 33 se observa que el 100% de los encuestados dicen que si existe un procedimiento formal para dar de alta y de baja a los usuarios

Figura 34

Pregunta 10

¿Se ha implementado un proceso formal de aprovisionamiento de accesos a los usuarios para asignar o revocar derechos de acceso a todos los tipos de usuarios y para todos los sistemas y servicios?

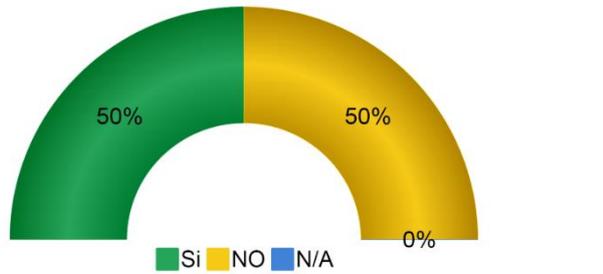


En la Figura 34 se observa que el 50% de los encuestados que si existe un proceso formal para asignar y revocar derechos de acceso a los usuarios y el otro 50% dice que no.

Figura 35

Pregunta 11

¿La asignación y uso de derechos de acceso con privilegios especiales son restringidos y controlados?

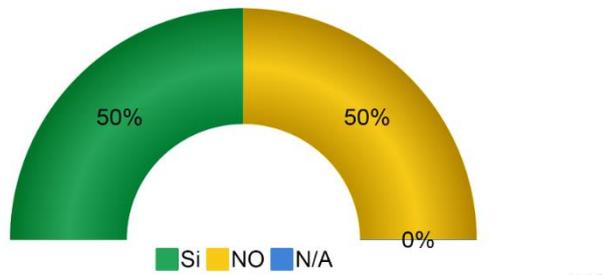


En la Figura 35 se observa que el 50% de los encuestados dice que si son controlados la asignación de derechos especiales y el 50% restante dice que no.

Figura 36

Pregunta 12

¿ La asignación de información confidencial para la autenticación es supervisada mediante un proceso de gestión controlado?

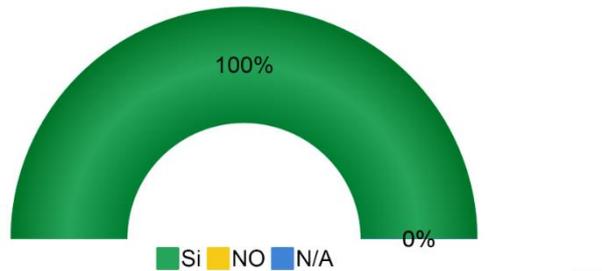


En la Figura 36 se observa que el 50% de los encuestados si cuenta con asignación de información confidencial para la autenticación supervisada y el 50 % restante dice que no.

Figura 37

Pregunta 13

¿Se exige a los usuarios el uso de las buenas prácticas de seguridad en el uso de información confidencial para la autenticación?

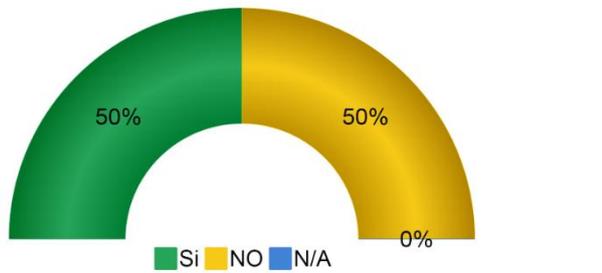


En la Figura 37 se observa que el 100% de los encuestados dice que si se exige a los usuarios buenas prácticas de seguridad en uso de la información confidencial.

Figura 38

Pregunta 14

¿Se restringe el acceso a los usuarios y al personal de mantenimiento de sistemas y aplicaciones, en relación a la política de control de accesos definida?

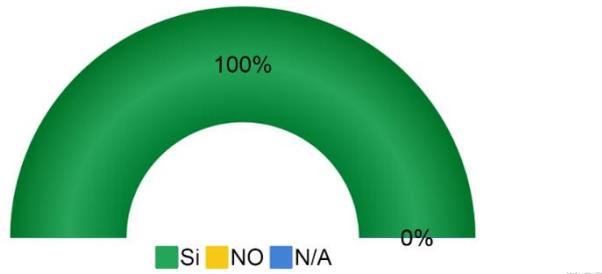


En la Figura 38 se observa que el 50% de los encuestados afirma que, si se restringe a los usuarios al personal de mantenimiento del sistemas y aplicaciones, el 50% restante afirma que no.

Figura 39

Pregunta 15

¿Tienen procedimientos seguros de inicio de sesión?

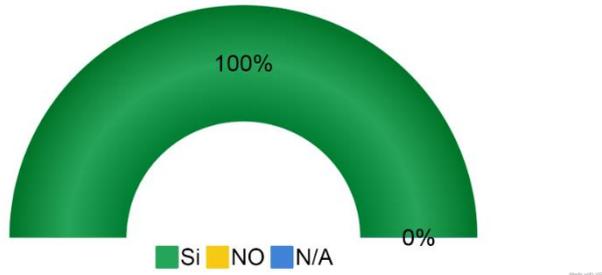


En la Figura 39 el 100% de los encuestados concuerda con que si se tiene procedimientos seguros de inicio de sesión.

Figura 40

Pregunta 16

¿Tienen un manejo adecuado de gestión de contraseñas de usuario?

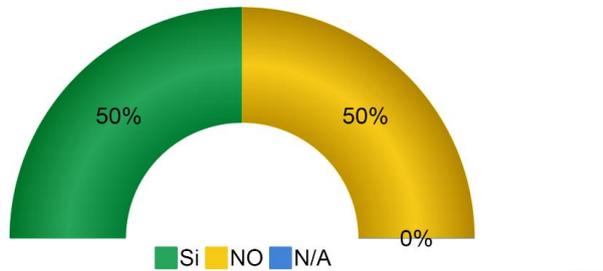


En la Figura 40 se observa que el 100% de los encuestados dicen que si se tiene un manejo adecuado de gestión de contraseñas de usuario.

Figura 41

Pregunta 17

¿Se dispone de herramientas de administración de sistemas y estas son restringidas y estrechamente controladas?

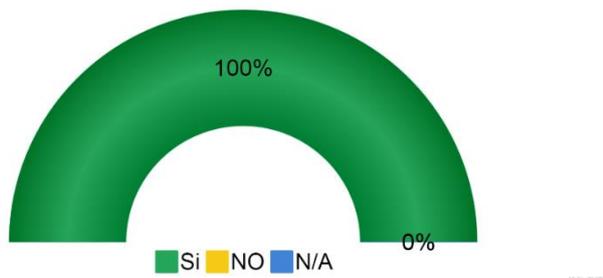


En la Figura 41 se observa el 50% de los encuestados si dispone de herramientas de administración del sistema y estas son restringidas y el 50% restante afirma que no.

Figura 42

Pregunta 18

¿Existe una restricción del acceso al código fuente de las aplicaciones software?

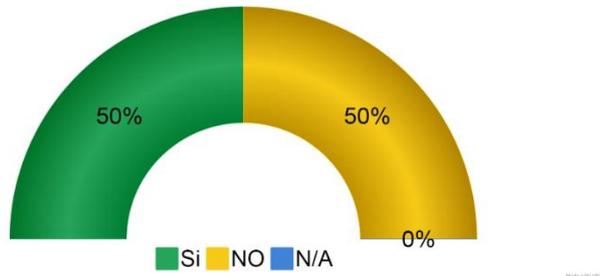


En la Figura 42 se observa que el 100% de los encuestados afirma que si existe restricciones de acceso al código fuente de aplicaciones de software.

Figura 43

Pregunta 19

¿Se protege contra posibles alteraciones y accesos no autorizados la información de los registros?

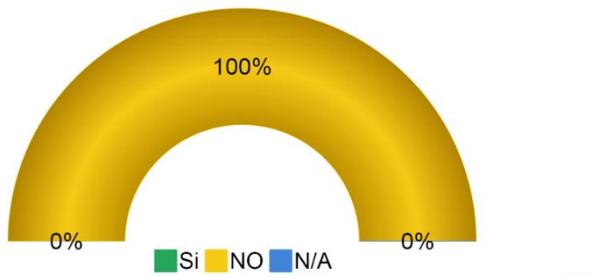


En la Figura 43 se observa que el 50% de los encuestados si se protege contra posibles alteraciones y accesos no autorizados la información de los registros, mientras que el 50% restante afirma que no.

Figura 44

Pregunta 20

¿Se registran las actividades del administrador, del operador del sistema y los registros asociados son revisados de manera regular?

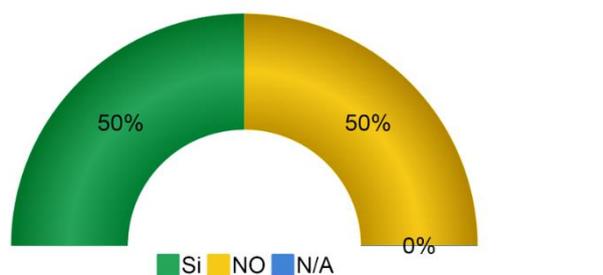


En la Figura 44 se observa que el 100% de los encuestados dicen que no se registran las actividades del administrador, del operador del sistema y los registros asociados son revisados de manera regular.

Figura 45

Pregunta 21

¿Se han planificado y acordado los requisitos y las actividades de auditoría que involucran la verificación de los sistemas operacionales?



En la Figura 45 el 50% de los encuestados dice que si se han planificado y acordado requisitos y las actividades de auditoría que involucran la verificación de los sistemas operacionales el 50% restante afirma que no se realizan.

Tabulación de entrevista

Tema: Estudio de Técnicas Avanzadas de Análisis de Datos para el Proceso de Auditoría Informática en el Módulo del Sistema Académico de la Universidad Técnica del Norte.

Objetivo: Verificar la situación actual de la calidad de los datos en el módulo de gestión de notas y calificaciones del sistema académico de la Universidad Técnica del Norte en base a la norma ISO/IEC 25012:2008

Entrevista realizada a: Ing. Jorge Caraguay director del Departamento de Desarrollo Tecnológico.

1. ¿En la institución que usted se encuentra laborando actualmente qué tipo de auditoría se realiza?

Utilizamos los módulos de auditoría de Oracle en toda la funcionalidad en la base de datos

- 2. ¿Si su respuesta a la pregunta anterior fue afirmativa, indique si la auditoría fue realizada por auditores internos o externos y qué áreas del departamento fueron auditadas?**

Se realiza por parte interna de la base de datos no se realiza auditorias de auditores interno o externo

- 3. ¿Cada que tiempo y bajo que motivos se realizan auditorias informáticas en el departamento de tecnologías de su institución?**

Es una auditoria permanente, auditorias formales al desarrollo de tecnología nunca se han realizado

- 4. ¿Dentro del proceso de auditoría se utilizaron técnicas de avanzadas de análisis de datos para identificar patrones en los datos, Cuáles fueron las técnicas aplicadas?**

El proceso de auditoría de Oracle se actualiza permanentemente y utiliza la normativa de modelo actualizo vigentes

- 5. ¿Cómo se controla y se mide la confidencialidad de los datos del sistema académico de la institución?**

Se utiliza las métricas de Oracle y las políticas internas de roles de usuarios, se entregan acuerdos de nivel de servicios a cada uno de los servicios según su rol

- 6. ¿Cómo se garantiza la integridad y la confiabilidad de los datos a lo largo de todo el ciclo de vida del sistema académico?**

Bajo las políticas internas de gestión de datos

- 7. ¿Puede describir quiénes tienen acceso y la autoridad para subir y modificar notas y calificaciones?**

1.- El docente en los periodos definidos

2.- quienes fugen de secretarios de las carreras bajo un documento emitido por consejo directivo de cada unidad académico, en este caso existe un campo donde se sube la evidencia de la justificación

8. ¿Cuál es el proceso que se debe llevar para modificar notas en el módulo de gestión de notas y calificaciones?

Se establece que para modificar notas el argumento o justificación debe ser conocido a través de un trámite por el consejo directivo, quienes evalúan de acuerdo a las normas vigentes si procede o no, en caso de que proceda se emite la resolución respectiva dirigida a la carrera para que la secretaria modifique la nota de acuerdo al caso, esta resolución ira con copia al DDTI para conocimiento y marca de evidencia

9. ¿Cuál es el tiempo administrado por el sistema para subir y modificar notas?

Existen los tiempos definidos en el calendario académico aprobados por el HCU en el que los docentes pueden subir las notas oficialmente; existen ocasiones, en que el HCU amplía o modifica estos periodos, esto puede generar confusiones por lo que existe un campo en el que insertan en el core de la base de datos las evidencias de estas modificaciones

10. ¿Bajo qué motivos se permite el modificar notas después de culminada la fecha de asentamiento de notas?

Es un análisis que está en potestad del consejo directivo y consejo universitario quienes analizan la normativa vigente

11. ¿Qué prácticas o técnicas implementa para asegurar la precisión de los datos del sistema académico?

La modificación de los datos es en línea por lo que no existe riesgo para la precisión de los datos en la normativa vigente

12. ¿Según su experiencia cuales cree que son los protocolos o procedimientos establecidos para resolver inconsistencias de imprecisión identificadas en el sistema académico?

No sé a detectado imprecisión en las notas, todo depende del usuario que digite las notas, posibles errores podrían venir de uso inadecuado de las contraseñas

13. ¿Se lleva a cabo alguna auditoría o revisión periódica para verificar la trazabilidad del sistema académico?

La auditoría permanente de Oracle mensualmente genera un reporte

14. ¿Se realiza un seguimiento y registro de todas las modificaciones realizadas en el sistema académico? ¿Cómo se documentan estas modificaciones?

Existen campos dentro del base de datos que sirven para registrar las resoluciones para ser justificadas

15. ¿Existe algún mecanismo para rastrear la relación entre los datos ingresados en el sistema académico y los resultados generados a partir de ellos?

Se registra usuario e ip

16. ¿Se ofrecen capacitaciones o sesiones de formación para los usuarios del sistema académico, con el fin de mejorar su comprensión y aprovechamiento de sus características?

Se da cursos de inducción y asistencia permanente por parte de los técnicos, también se elaboran manuales de usuario

17. ¿Se recopilan y analizan los comentarios y sugerencias de los usuarios para mejorar la comprensibilidad del sistema académico en futuras iteraciones o versiones?

La retroalimentación es permanente con la fuente de los roles de usuario de secretarios

Análisis

Durante la encuesta se pudo conocer que sistema académico universitario cuenta con políticas de seguridad para garantizar la confidencialidad y brindada capacitación a los usuarios y correcto uso de contraseñas, también se pudo conocer que las únicas personas para modificar notas es docente y secretario de carrera bajo justificación y aprobación de consejo directivo. De igual manera HCU es el encargado de establecer y modificar los tiempos del calendario académico

Análisis de clúster

Las Figuras 14 - 24 muestran los 11 clústeres de los primeros 20 registros de notas de las 5 facultades, en los que se puede visualizar que, los puntajes de calificación bajos reducen la cantidad de datos de agrupación en cada clúster, lo que sugiere que estos puntajes fueron alterados, se pasa comparar las notas finales con los registros modificados para comprobar y verificar si los datos son precisos.

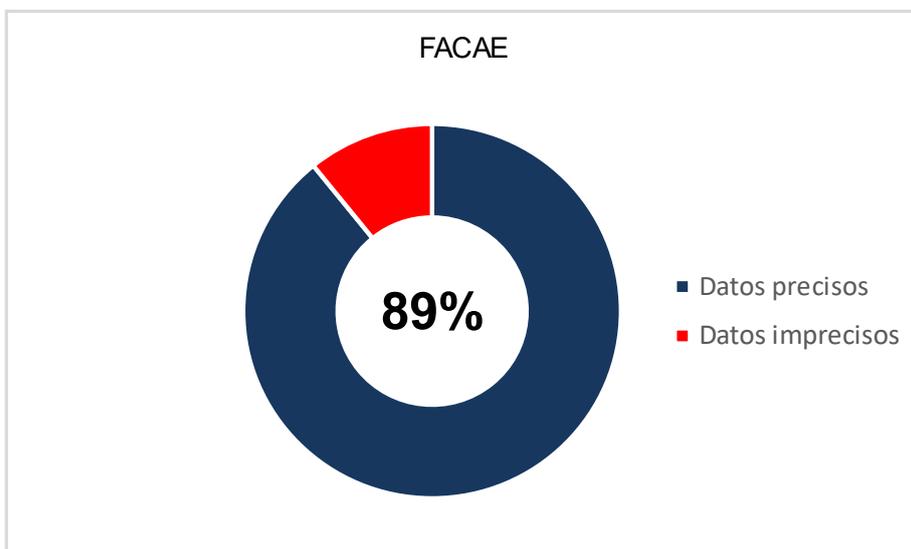
Análisis descriptivo

Precisión. A continuación, se verifico si las notas modificadas de los 62 registros realizados coinciden con las notas finales NOTA1 y NOTA2, también se revisó

las fechas de registro para comprobar si las modificaciones se efectuaron dentro del cronograma planificado del periodo.

Figura 46

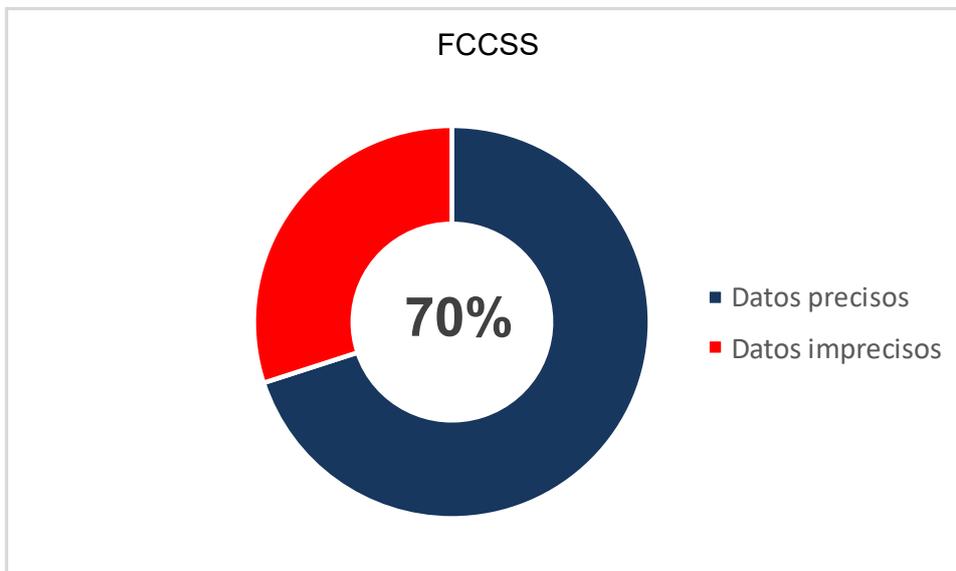
Precisión de datos FACAE



En la Figura 46 se puede observar que el 11% que equivale a 1.517 de los datos de la facultad FACAE son diferentes, es decir, las notas modificadas no son precisas y no coinciden con las notas finales y un alto porcentaje del 89% que equivale a 12.493 de los datos modificados coinciden con las notas finales.

Figura 47

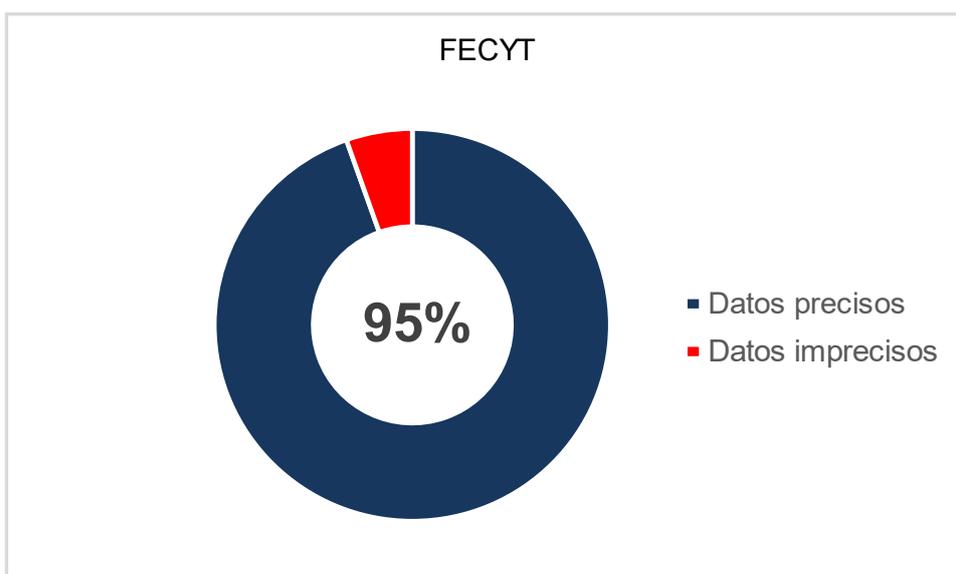
Precisión de datos FCCSS



En la Figura 47 se observa que el 30% equivalente a 1.969 de los datos de la facultad de FCCSS, no son precisos, las notas modificadas en los registros no coinciden con las notas finales a diferencia de un alto porcentaje del 70% de los datos equivalente a 4.592 notas que sí coinciden las notas modificadas con las notas finales.

Figura 48

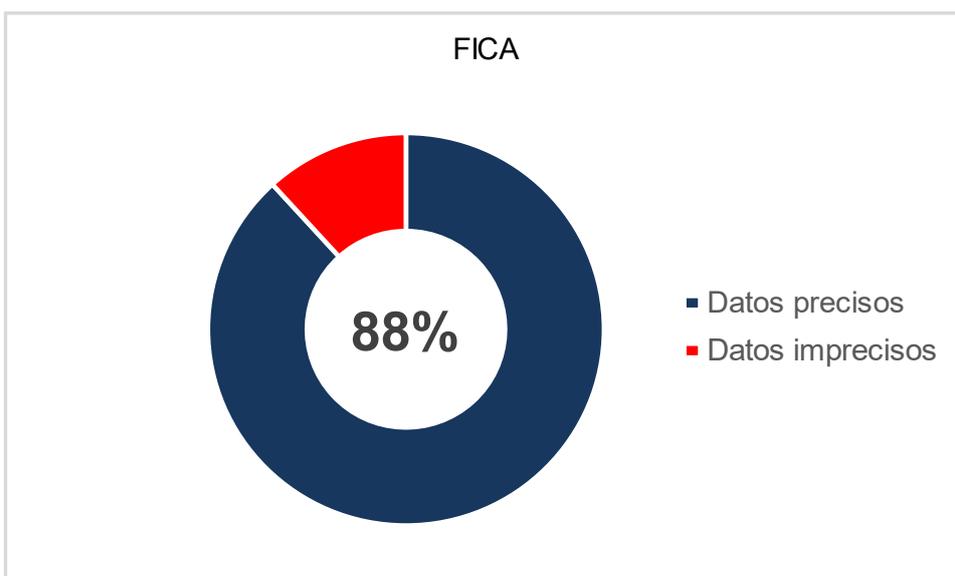
Precisión de datos FECYT



En la Figura 48 referente a la facultad FECYT, se observa un porcentaje muy bajo del 5% equivalente a 1.246 datos que no son precisos, las modificaciones de notas no coinciden con las notas finales, el otro 95% de las notas modificadas son precisas y coinciden con las notas finales.

Figura 49

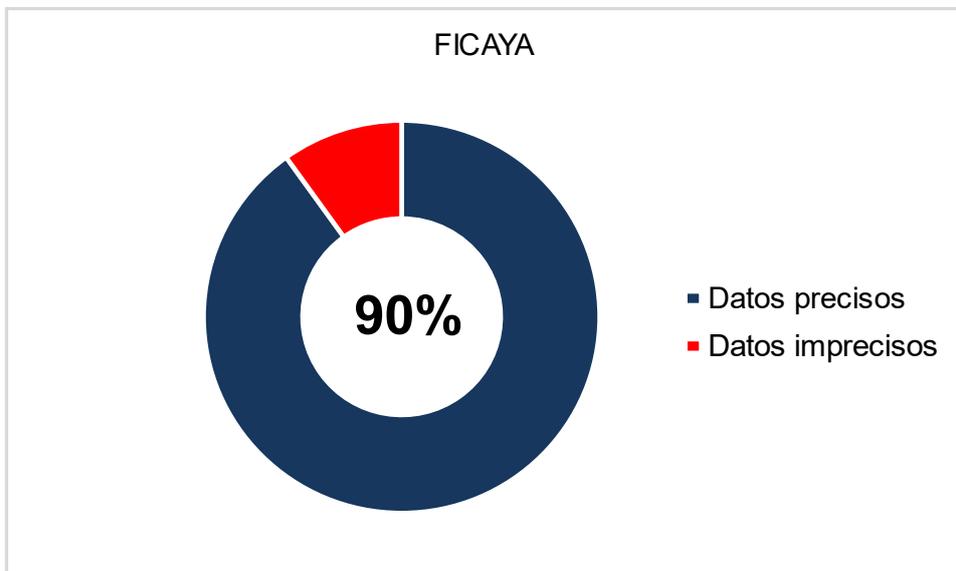
Precisión de datos FICA



En la Figura 49 que corresponde a la facultad FICA, se puede observar que el 12% de los datos referente a 876 notas no son precisas, las notas modificadas no coinciden con las notas finales, el 88% restante que equivale a 6.560 notas son precisas, las notas modificadas coinciden con las notas finales

Figura 50

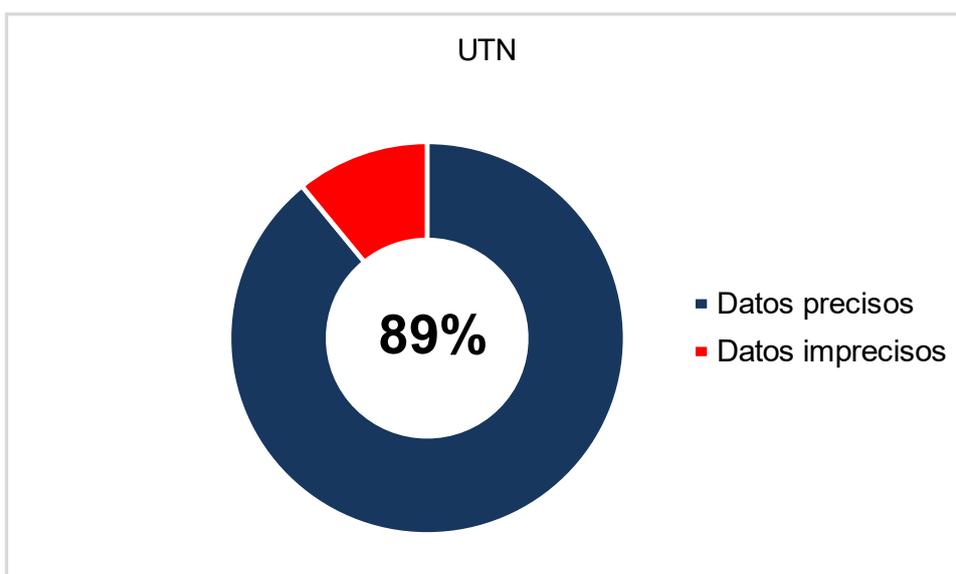
Precisión de datos FICAYA



En la Figura 50 se puede observar que el 10% que equivale a 488 de los datos de la facultad FICAYA son diferentes, es decir, las notas modificadas no son precisas y no coinciden con las notas finales y un alto porcentaje del 90% que equivale a 4438 de los datos modificados coinciden con las notas finales.

Figura 51

Precisión de datos UTN

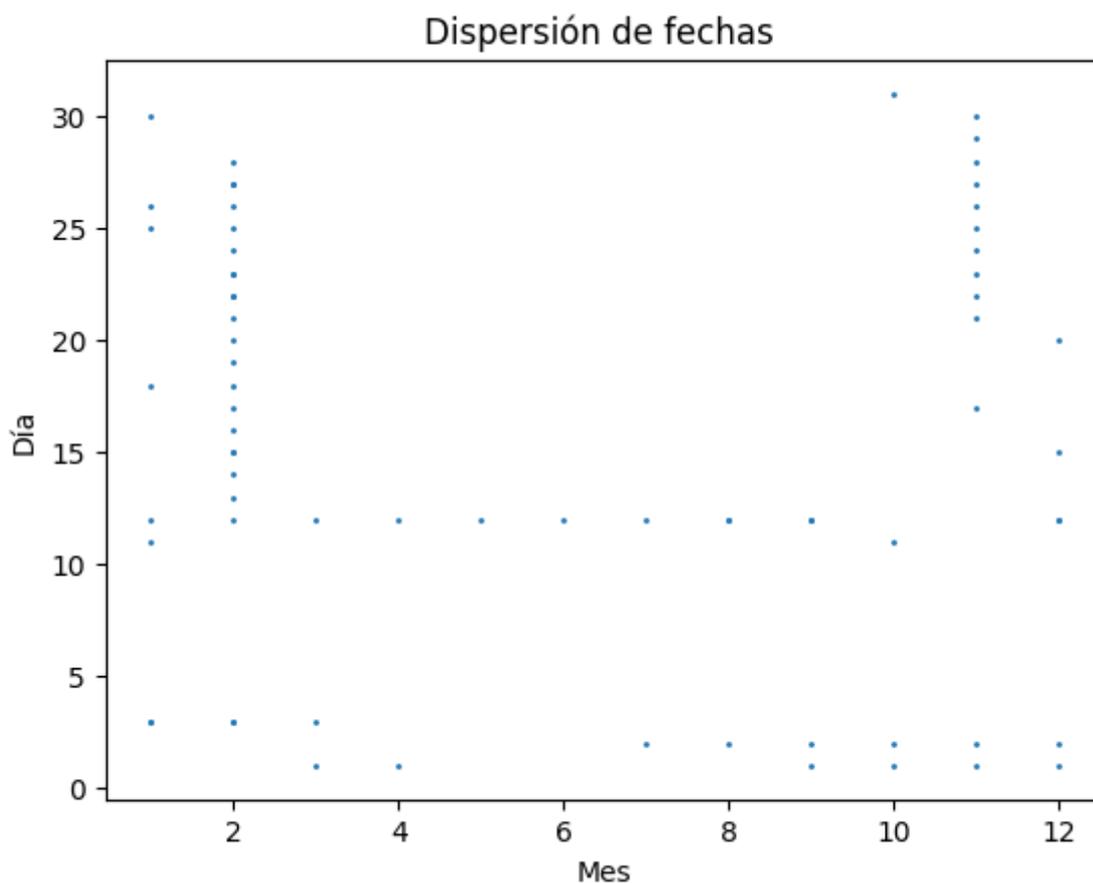


En datos generales del periodo SEP2022-FEB2023 se analizaron un total de 55.974 datos, por lo que se puede observar en la Figura 51 el 11% de los datos referente a 6096 datos no son precisos, las notas modificadas no coinciden con las notas finales, el 89% restante que equivale a 49.878 de las notas son precisas, las notas modificadas coinciden con las notas finales.

También se realizó un análisis para comprobar si se cumple la modificación de notas en las fechas establecidas en el calendario académico y así verificar la precisión de los datos, en la Figura 52 se observa que existe modificaciones fuera de las fechas establecidas la primera parcial se realiza del 28 de noviembre al 04 de diciembre del 2022 y la segunda parcial del 13 de febrero al 19 de febrero del 2023.

Figura 52

Dispersión de Fechas



Confidencialidad. De acuerdo con la norma ISO/IEC 25012:2008 se encuentra que los archivos de modificación de notas no fueron modificados y alterados por personas externas y cumplen con la característica de confidencialidad.

Trazabilidad. De acuerdo con la norma ISO/IEC 25012:2008 los datos cuentan con fácil acceso para ser auditados y monitorear las acciones de los usuarios del sistema por lo que cumple con esta característica.

Comprensibilidad. De acuerdo con la norma ISO/IEC 25012:2008 los atributos de los datos cumplen este requisito, lo que los hace fáciles de comprender cuando se analizan en grandes conjuntos de datos y se aplican técnicas avanzadas de análisis de datos.

2.3.4. Conclusiones preliminares

En la Tabla 22 se muestran los hallazgos encontrados durante la auditoría en base a la norma ISO/IEC 25012:2008, las evidencias de datos no pueden ser publicas al firmarse la carta de confidencialidad con la institución.

Tabla 23

Tabla Hallazgos

Tabla de Hallazgos			
N°	Hallazgos	Descripción	Evidencia
1	Imprecisión de los datos	La imprecisión de la modificación de las notas de las diferentes facultas está en FACA E 11%, FICAYA 10%, FICA 12%, FECYT 5%, FCCSS 30%	Figuras 46 - 51
2	Modificación de notas fuera del calendario académico	De acuerdo con el calendario académico la primera parcial se realiza del 28 de noviembre al 04 de diciembre del 2022 y la	Figura 52

segunda parcial del 13 de febrero
al 19 de febrero del 2023 y durante
el análisis realizado se observa
modificaciones fuera de las fechas
establecidas.

En el primer hallazgo encontrado puede deberse a diferentes factores, es fundamental abordar los problemas relacionados con la imprecisión de las notas en el sistema informático académico de inmediato, ya que puede plantear serias dificultades tanto para los estudiantes como para los docentes afectando la transparencia y credibilidad del sistema.

En el segundo hallazgo encontrado puede comprometer las políticas de control interno con la libertad de modificar o subir notas fuera del calendario académico establecido, dando posibilidad al fraude académico, poniendo en peligro la integridad y el cumplimiento de los registros académicos.

2.4. Comunicación De Resultados (FASE III)

2.4.1. *Elaboración Del Informe Final Preliminar*

El informe preliminar se efectuó de acuerdo se indica en el plan preliminar y se realizaron las recomendaciones de acuerdo con las evidencias encontradas por el equipo auditor en el reporte de modificación de notas del sistema de académico de la Universidad técnica del norte.

2.4.2. Informe Final Preliminar

Introducción

La UTN (Universidad Técnica del Norte) como referente de la región norte del Ecuador desarrolla sus aplicaciones informáticas y las integra al SIIU (Sistema Informático Integrado de las Universidades), la cantidad de información aumenta

exponencialmente con el tiempo, la seguridad de los sistemas informáticos también han aumento siendo expuesta a ataques y en efecto a la vulnerabilidad de los datos.

Datos de los auditores

Rol	Nombres y apellidos
Auditor Líder	Daisy Elizabeth Imbaquingo Esparza
Auditor Junior	Francisco Javier Álvarez Osorio

Alcance

Para realizar la auditoría del sistema académico de la Universidad del Norte en el módulo de gestión de notas y calificaciones, se toma en cuenta la norma ISO/IEC 25012:2008 para comprobar la calidad de los datos inherente y dependiente del sistema como la confidencialidad, precisión, trazabilidad y comprensibilidad, aplicando técnica avanzada de análisis de datos para descubrimiento de patrones en el registro de modificación de notas, de las facultades FACAE, FCCSS, FECYT, FICA, FICAYA, del periodo SEP2022-FEB2023.

Objetivos

Objetivo General

Realizar una auditoría informática de los datos del sistema académico de la UTN aplicando técnicas avanzadas de análisis de datos para determinar su calidad.

Objetivos Específicos

- Recolección y limpieza de datos del módulo de gestión de notas y calificaciones.
- Analizar los datos con técnicas de minería de datos.
- Elaborar un informe técnico con recomendaciones en base a los resultados obtenidos con la norma ISO/IEC 25012:2008

Metodología, herramientas y estándar utilizados

Se utiliza la metodología CRISP-DM denominada así por sus siglas en inglés (Cross Industry Standard Process for Data Mining) es una metodología para minería de datos

que cuenta con ciclo de vida de 6 fases las cuales se indican en la Figura 11, además se utiliza como herramienta de procesamiento el lenguaje de programación Python en la plataforma Google Colab para implementar el algoritmo de clúster ing K-menens con el fin de detectar patrones en el conjunto de datos y ayudar a identificar anomalías bajo el estándar ISO/IEC 25012:2008 de calidad de datos.

Cronograma

Se ha programado que la auditoría se la realizará en un tiempo estimado de 12 semanas, las cuales se contará desde el día 10 de abril de 2023.																	
ACTIVIDADES	RESULTADOS POR OBJETIVO / ACTIVIDAD	CRONOGRAMA DE ACTIVIDADES												% de avance / actividad	% de cumplimiento / objetivo	% avance TOTAL	
		SEMANAS 2023															
		15	16	17	18	19	20	21	22	23	24	25	26				
Contacto con el cliente	brindar información	█													100%	100%	100%
Estudio Inicial del entorno a auditar	Conocer la institución	█													100%	100%	100%
Definir el equipo auditor	Asesoramiento de auditoría	█													100%	100%	100%
Elaboración del Plan Preliminar	Informe preliminar	█	█	█											100%	100%	100%
Elaboración de propuesta de auditoría	Recomendaciones a la institución		█	█											100%	100%	100%
Contrato	Establecer cumplimientos		█	█											100%	100%	100%
Ejecución de la auditoría	Verificar situación actual			█	█	█	█	█	█						100%	100%	100%
Validación de la auditoría	Evaluación de auditoría								█	█	█				100%	100%	100%
Seguimiento	Informe final										█	█	█		100%	100%	100%

Checklist

En el área de confidencialidad se garantizó la correcta aplicación de las políticas de seguridad, gestión de roles, capacitación de usuarios, a través de la investigación de campo que incluyó entrevista y encuesta.

La precisión se comprobó que las notas modificadas coincidan con las notas finales por lo que se encontró irregularidades que no cumple con esta característica.

Sobre la base de los reportes de notas y la investigación de campo, se pudo confirmar que la trazabilidad de los datos es útil para auditar el cumplimiento con un historial de modificaciones de notas, registro de actividades a través de usuarios e IP, y auditorías periódicas a la base de datos.

Finalmente, la comprensibilidad de los datos se pudo comprobar con el reporte de modificación de notas al ser fácil de utilizar y comprender con las técnicas de minería de datos para identificar patrones.

Características	Checklist
Confidencialidad	✓
Precisión	○
Trazabilidad	✓
Comprensibilidad	✓

Hallazgos

En la Tabla 21 se muestran los hallazgos encontrados durante la auditoría en base a la norma ISO/IEC 25012:2008, las evidencias de datos no pueden ser publicas al firmarse la carta de confidencialidad con la institución.

Tabla de Hallazgos			
N°	Hallazgos	Descripción	Evidencia
1	Imprecisión de los datos	La imprecisión de la modificación de las notas de las diferentes facultades está en FACAE 11%, FICAYA 10%, FICA 12%, FECYT 5%, FCCSS 30%	Figuras 46 - 51

<p>2</p> <p>Modificación de notas fuera del calendario académico</p>	<p>De acuerdo con el calendario académico la primera parcial se realiza del 28 de noviembre al 04 de diciembre del 2022 y la segunda parcial del 13 de febrero al 19 de febrero del 2023 y durante el análisis realizado se observa modificaciones fuera de las fechas establecidas.</p>	<p>Figura 52</p>
--	--	------------------

Matriz de riesgos

RIESGOS		
INHERENTES	DE CONTROL	DE DETECCIÓN
1. Factores externos.		1. Habilidades blandas
2. Objetivos y		del equipo auditor.
estrategias.		2. Objetivos y alcance
3. Control interno.	1. Supervisión.	claros.
4. Disponibilidad de	2. Comunicación.	3. Aptitud y actualidad
tiempo en los	3. Mecanismo de control	del equipo auditor.
encargados de la	de gestión u operación.	4. Enfermedad o
IES.	4. Ambiente o entorno de	percances en equipo
5. Falta de	control.	auditor.
documentación.	5. Evaluación de riesgos.	5. Conocimiento previo
6. Causas de		de la IES para auditar.
incorrecciones.		6. Conocimiento de
7. Información		técnicas de auditoría.
fraudulenta		

Conclusiones

En el primer hallazgo encontrado puede deberse a diferentes factores, es fundamental abordar los problemas relacionados con la imprecisión de las notas en el sistema informático académico al no coincidir las notas modificadas con las notas finales, ya que puede plantear serias dificultades tanto para los estudiantes como para los docentes afectando la transparencia y credibilidad del sistema.

En el segundo hallazgo encontrado puede comprometer las políticas de control interno con la libertad de modificar o subir notas fuera del calendario académico establecido, dando posibilidad al fraude académico, poniendo en peligro la integridad y el cumplimiento de los registros académicos.

CAPÍTULO III

Validación

3.1. Validación (FASE IV)

Para garantizar la transparencia y confiabilidad de los resultados de la auditoría, la validación de la auditoría es un proceso esencial. A través de la validación se evaluarán parámetros de calidad, seguridad de la información y cumplimiento.

3.1.1. Calidad

Se utilizan tres factores (humano, técnico y contextual o ambiental) para evaluar la calidad de la auditoría. La Tabla 23 enumera las preguntas para cada aspecto.

Tabla 24

Preguntas de evaluación de calidad

FACTOR DE EVALUACIÓN	MÉTRICA DE EVALUACIÓN
	El equipo auditor procuró que el cliente participe en todo el proceso de auditoría
	El equipo auditor obtuvo la conformidad del cliente acerca de las actividades desarrolladas
Factor Humano	El personal que realiza la auditoría tenía competencias necesarias para realizar su trabajo
	El auditor estaba seguro de sí mismo y de su trabajo
	El equipo auditor conservó su independencia en apariencia y acción

El equipo auditor se centró en los hechos

El equipo auditor recibió apoyo para lograr las metas

El equipo auditor demuestra esfuerzo al realizar la auditoría

El auditor se preocupaba por su formación y actualización continua

El auditor contaba con certificaciones nacionales e internacionales en el área de auditoría y auditoría informática

Los miembros del equipo auditor demostraron conocimiento en seguridad de la información y procesamiento de datos

Las diferencias con el cliente fueron tratadas de forma oportuna, profesional y objetiva

El auditor vinculó expertos como apoyo en el proceso de auditoría para obtener resultados y recomendaciones para el cliente

El equipo auditor usó plantillas y formularios para documentar

Los hallazgos y conclusiones de la auditoría fueron un reflejo exacto de los hechos reales del proceso auditado

Factor Técnico

Los resultados de la auditoría fueron respaldados y documentados con las evidencias recopiladas al auditar

Los recursos para la auditoría fueron asignados de acuerdo con la importancia y complejidad de la auditoría

El sistema, proceso u objeto auditado tenía importancia para la organización

En el alcance se abordaron todos los elementos necesarios para auditar exitosamente

La ejecución de la auditoría cumplió con los elementos acordados en el alcance

El modelo de evaluación de riesgos fue comprensible

El plan de auditoría tomó en cuenta los riesgos relacionados con el cliente

El proceso de auditoría se desarrolló con exactitud y precisión

El informe de auditoría fue claro y conciso con sus resultados

El alcance, hallazgos y recomendaciones han sido entendibles para cualquier persona que haga uso del informe de auditoría

La auditoría se ejecutó bajo las políticas, estándares, manuales, directrices y prácticas de auditoría informática

Las listas de verificación estuvieron completas, aprobadas y documentadas

El trabajo de campo fue revisado por un experto

La información y resultados de anteriores auditorías estuvieron disponibles para revisión

Los objetivos y el alcance de la auditoría fueron especificados adecuadamente

Los miembros del equipo auditor tenían una comprensión clara y coherente del plan de auditoría

	El presupuesto y cronograma de auditoría se establecieron de manera adecuada
	Se evaluaron los requisitos de personal y equipos asignados para la auditoría
	El plan de auditoría fue elaborado, revisado y aprobado por los supervisores, responsables de la organización y miembros del equipo auditor
	El equipo auditor utilizó una metodología de auditoría informática para planificar, gestionar y desarrollar la auditoría
	El auditor promovió a través de sus informes una cultura organizacional basada en buenas prácticas de seguridad informática
	El equipo auditor tenía estrictos procedimientos de control de calidad
	El líder del equipo auditor estuvo comprometido con el sistema de control de calidad
Factor Contextual o del Entorno	La normativa y regulaciones emitidas por organismos de control fueron reflejadas en el plan de auditoría
	El equipo auditor conocía la información relevante de leyes y regulaciones que puedan tener un impacto significativo en los objetivos de la auditoría
	Se aplicaron medidas disciplinarias en caso de incumplir con el plan de auditoría o la normativa legal regulatoria vigente
	El costo de la auditoría estuvo de acuerdo con la complejidad y las actividades desarrolladas

En la Tabla 24 se muestra los resultados obtenidos por cada evaluador en ámbitos de calidad.

Tabla 25*Resultados de Calidad*

Evaluador	Porcentaje
Evaluador 1	70%
Evaluador 2	67%

Tras la validación se obtiene un promedio de calidad general del 78,5%, que se categoriza como calidad media.

3.1.2. Seguridad de la información

En el área de seguridad se evalúan los factores de confidencialidad, integridad y disponibilidad las cuales se observan en la Tabla 25.

Tabla 26*Preguntas de evaluación de seguridad*

FACTOR DE EVALUACIÓN	MÉTRICA DE EVALUACIÓN
Confidencialidad	Se aplican políticas para la seguridad de la información dentro de la institución
	Las políticas y procedimientos en seguridad de la información dentro de la institución se actualizan periódicamente
	Las responsabilidades en la seguridad de la información son delegadas, documentadas y entregadas formalmente a todo el personal de la institución, según su cargo
	Se aplican políticas y acciones de seguridad de la información sensible de la institución
	Se actualiza y aplica las políticas de acceso a la información en base a los roles de usuario existentes

Se dispone de una acreditación en seguridad de la información para todos sus sistemas informáticos

Se aplican procedimientos documentados para seguir en caso de incidentes de seguridad

Se realizan auditorías de cumplimiento de seguridad de la información

Se aplican políticas de gestión de contraseñas para los usuarios finales de la institución

Se identifican a los usuarios que acceden a la red y las acciones que ejecutan

Se aplica un control de acceso a la infraestructura y servicios de TI de la institución

Se capacita e involucra a usuarios, colaboradores y personal en los temas de seguridad de la información

Integridad

Se realiza análisis de vulnerabilidades de los servicios web de la institución

Se aplican planes de monitoreo y gestión de impacto de incidentes de seguridad en la institución

Se actualiza y documenta el inventario de todos los activos de TI

Se dispone de aplicaciones para proteger de software malicioso a todas sus soluciones informáticas

Disponibilidad

Se realizan copias de seguridad de la información

Se monitorean las actividades desarrolladas por los usuarios

En la Tabla 26 se muestra los resultados obtenidos por cada evaluador en ámbitos de seguridad.

Tabla 27

Resultados de seguridad

Evaluador	Porcentaje
Evaluador 1	82%
Evaluador 2	76 %

Después de realizar la validación se obtiene un promedio de calidad general de 84%, que se categoriza cómo seguridad media.

3.1.3. Cumplimiento

En cumplimiento se evalúa la planificación de la auditoría, ejecución de la auditoría, comunicación de resultados, validación de la auditoría, seguimiento de la auditoría, las que se puede observar en la Tabla 27.

Tabla 28

Preguntas de evaluación de cumplimiento

FASES DE LA AUDITORÍA	ACTIVIDAD
	Establecer lugar y fecha de encuentro con el cliente
Planificación de la auditoría	
	Registrar reunión inicial con el cliente

Identificar las partes interesadas y responsables para la auditoría

Comprender el contexto externo del entorno a auditar

Comprender el contexto interno de la institución a auditar

Comprender estrategias y prioridades de la institución

Determinar objetivos de la auditoría en base al estudio del entorno

Determinar el riesgo relevante del proceso de auditoría

Definir los límites organizacionales de la auditoría

Identificar miembros del equipo auditor

Seleccionar miembros del equipo auditor

Definir roles acordes a los conocimientos y habilidades de los miembros del equipo auditor

Documentar plan preliminar

Determinar antecedentes

Determinar objetivos específicos de la auditoría

Determinar el alcance de la auditoría

Determinar recursos para la auditoría

Determinar cronograma de la auditoría

Determinar costos de la auditoría

Documentar los riesgos de la auditoría

Presentar plan preliminar al cliente

Acordar términos y condiciones de la auditoría

Elaborar la propuesta de auditoría

Presentar la propuesta de la auditoría

	Determinar cláusulas del contrato
	Elaborar el contrato
	Firmar el contrato
	Diseñar y documentar el plan de trabajo
	Definir instrumentos para la investigación de campo
	Elaborar instrumentos de investigación de campo determinados en el plan de trabajo
Ejecución de la auditoría	Aplicar técnicas e instrumentos de auditoría
	Análisis y síntesis de información recopilada
	Definir hallazgos en base a la información y evidencias recopiladas
	Elaborar informe final preliminar
Comunicación de resultados	Agendar fecha para presentación y discusión del informe final preliminar

	Presentar informe final preliminar para discusión y validación
	Elaborar informe final
	Agendar fecha para presentación de informe final
	Presentar informe final y documentos resultantes de la auditoría
	Cerrar el contrato
	Retirar garantías
Validación de la auditoría	Aplicar evaluación de calidad de resultados de auditoría
	Aplicar evaluación de seguridad de la información en las IES
	Aplicar evaluación de cumplimiento de actividades de la auditoría
	Utilización de Data Mining
Seguimiento de la auditoría	Consulta directa al auditado
	Verificación del informe final de auditoría

En la Tabla 28 se puede observar los resultados obtenidos en ámbitos de cumplimiento

Tabla 29

Resultados de cumplimiento

Evaluador	Porcentaje
Evaluador 1	83%
Evaluador 2	89 %

Después de realizar la validación se obtiene un promedio de calidad general de 86%, que se categoriza como nivel de cumplimiento alto.

3.2. Seguimiento (FASE V)

3.2.1. Informe Final de Auditoría

Se revisa el informe final con el progreso de la implementación de acciones correctivas a las partes interesadas relevantes. Esto se realiza en un tiempo prudente de 6 meses a 12 meses para brindar transparencia y rendición de cuentas para el progreso de la mejora.

3.2.2. Contacto con el Cliente

Se confirma al final de la fase de seguimiento que todas las acciones correctivas se han llevado a cabo con éxito y que se han obtenido los resultados deseados manteniendo a todos los interesados actualizados sobre el estado de la implementación por medio de comunicaciones escritas o reuniones de seguimiento con los interesados.

Conclusiones

Existen varias tecnologías y técnicas de procesamiento y análisis avanzado de datos que se pueden aplicar en auditorías, las características del conjunto de datos auditado dependen de la selección de la técnica a utilizar, en casos específicos estas pueden resultar de mucha ayuda como agrupación, asociación y correlación, es este caso de estudio la técnica que mejor se adapta es clúster ing al manejar variables cuantitativas sin etiquetar.

Al emplear una técnica de minería de datos para el procesamiento y análisis de datos, ayudó a identificar patrones de notas modificadas fuera del calendario académico establecido de una media del 15% de datos imprecisos, de acuerdo con la aplicación de la norma ISO/IEC 25012:2008.

El proceso de Auditoría informática fue evaluado con el método de auditoría informática para instituciones de educación superior (MAIIES) en índices de calidad con una media del 69%, seguridad de la información con una media del 79% y cumplimiento con una media del 86% por lo que muestra que la utilización de técnicas de procesamiento y análisis avanzado de datos en el proceso de auditoría informática es aceptable y puede ayudar a los auditores en el desarrollo de sus actividades.

Recomendaciones

Se debe realizar un análisis de las características de los datos y el objetivo de la auditoría para seleccionar una técnica que ayude a descubrir patrones y evidencias del conjunto de datos sumados a la experticia del auditor puedan brindar auditorías mucho más concretas.

Para realizar procesamiento y análisis de datos avanzados, el auditor necesita aprender nuevas habilidades y familiarizarse con nuevas herramientas, tecnologías de ciencia de datos, inteligencia artificial y big data, que son las fuentes de tecnologías complejas con una variedad de técnicas y algoritmos.

Se debe cumplir los tiempos establecidos teniendo en cuenta la privacidad de los datos, integridad y confidencialidad para brindar una auditoría de calidad realizando el seguimiento respectivo a las recomendaciones realizadas.

Referencias

- Ávalos Andrade, R., & Sainz-Grant, T. (1 de Abril de 2019). *Contaduría Pública*. Obtenido de Análisis de Datos en la Auditoría: <https://contaduriapublica.org.mx/2019/04/01/analisis-de-datos-en-la-auditoria/>
- Jifan, C., & Muhammad, T. (2021). Audit Data Analysis and Application Based on Correlation Analysis Algorithm. *Computational and Mathematical Methods in Medicine*, 1748-670, <https://doi.org/10.1155/2021/2059432>.
- Leventhal, B. (2010). An introduction to data mining and other techniques for advanced analytics. *Journal of Direct, Data and Digital Marketing Practice*, 137–153, <https://doi.org/10.1057/dddmp.2010.35>.
- Agresti, A. (2018). *Statistical Methods for the Social Sciences*. Pearson Education.
- Alexiou, S. (2017). Advanced Data Analytics for IT Auditors. *ISACA, Vol 6*, <https://www.isaca.org/es-es/resources/isaca-journal/issues/2016/volume-6/advanced-data-analytics-for-it-auditors>.
- Alteryx. (2019). *Herramienta de diagnóstico de K-centroides*. Obtenido de Agrupamiento predictivo: https://help.alteryx.com/2020.2/es/K-Centroids_Diagnostics.htm
- Arciniega Vera, A. J., & Ludeña Gonzalez, V. I. (2014). AUDITORÍA INFORMATICA BASADA EN RIESGOS DEL CORE BANCARIO DE UNA INSTITUCION FINANCIERA DE ACUERDO AL PLAN ANUAL 2014 DE AUDITORIA INTERNA. *ESCUELA POLITÉCNICA DEL EJÉRCITO*, 268.
- Arnau Sabatés, L., & Sala Roca, J. (2020). La revisión de la literatura científica: Pautas, procedimientos y criterios de calidad. *Universidad Autónoma de Barcelona*, 22.
- Astrid Shofi Dzihni, R. A. (2019). Business Process Analysis and Academic Information System Audit of Helpdesk Application using Genetic Algorithms a Process Mining Approach. *Procedia Computer Science*, 903-909.
- Beltran Camacho, N. S., & Bautista Cardenas, L. A. (2019). AUDITORÍA DE LA BASE DE DATOS DE SEGURIDAD SOCIAL DE LA EMPRESA INVERSIONES ALCABAMA S.A. *UNIVERSIDAD CATÓLICA DE COLOMBIA*, 1-126.
- Bernal, A. (2 de Enero de 2023). *Ciencia de datos*. Obtenido de GeoVictoria: <https://www.geovictoria.com/es-co/blog/ciencia-de-datos/>
- Blázquez Hernández, J. (17 de Mayo de 2018). *Auditoría & Co el portal de la auditoría*. Obtenido de La aplicación del Big Data y el Data Analytics en auditoría: <https://auditoria-audidores.com/articulos/articulo-auditoria-la-aplicacion-del-big-data-y-el-data-analytics-en-auditoria/>
- Böttcher, S., Vieluf, S., Bruno, E., & al, e. (2022). Data quality evaluation in wearable monitoring. *Scientific Reports* 12, 21412.
- Cevallos, H. V., Cuesta, J. D., & Montero, B. M. (2018). Gestión de la información en las instituciones de educación superior (IES) con base a la norma ISO 27001. *Zenodo*, 1-6.
- Chiriboga Toapanta, D. M. (2022). PROPUESTA DE ESTANDARIZACIÓN DEL PROCESO DE AUDITORÍA DE INFORMACIÓN PARA LAS INSTITUCIONES DE EDUCACIÓN SUPERIOR DE LA ZONA 1 DEL ECUADOR MEDIANTE NORMATIVAS INTERNACIONALES. *Universidad Tecnica del Norte*, 55.
- Demchenko, Y., Belloum, A., & Los, W. (2016). EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry. *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Luxembourg, Luxembourg*, 620-626.

- Duggal, N. (24 de Julio de 2023). *Big Data* . Obtenido de Simpli Learn: <https://www.simplilearn.com/what-is-data-processing-article>
- Espino Timó, C. (2017). Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso. *Universitat Oberta de Catalunya*, <http://hdl.handle.net/10609/59565>.
- Felix Krieger, P. D. (2018). Leveraging Big Data and Analytics for Auditing: Towards a Taxonomy. *DATA SCIENCE AND PREDICTIVE ANALYTICS*.
- Gandomi, A., Chen, F., & Abualigah, L. (2023). Big Data Analytics Using Artificial Intelligence. *Electronics* , 12.
- Gantz, S. D. (2014). *The Basics of IT Audit, Chapter 8 - IT Audit Processes*., <https://www.sciencedirect.com/science/article/pii/B9780124171596000080>: Syngress.
- Garavit, J. (2020). La Auditoría en la Gestión de Datos. *Universidad*, 5.
- Gartner. (2023). *Magic Quadrant*. Obtenido de Gartner: <https://www.gartner.es/es/metodologias/magic-quadrant>
- GHORPADE, Y., & MORE, R. (2015). *Information Security and Audit*. ISBN.
- González Mataix, P. (2018). Auditoría TI en la Asociación APSA. *Univerisdad de Alicante*, 64.
- Gonzalez Mataix, P., & Berná Martinez, J. V. (2018). Auditoría TI en la Asociación APSA. *Universidad de Alicante*, 206.
- González Silva, L. M., & Peña Rojas, W. A. (2022). Componente de software de análisis prescriptivo basado en las variables oceánicas y meteorológicas de Colombia. *Universidad Católica de Colombia*, 128.
- Grijalva Arriaga, P. K., Freire Avilés, V. A., & Real Avilés , K. P. (2018). Application of Data Mining Techniques for the Analysis of Academic Efficiency. *Revista Científica Hallazgos*, 21, <https://revistas.pucese.edu.ec/hallazgos21/article/view/222>.
- Hernández Obando, C. C. (2019). ESTUDIO DE LA SEGURIDAD EN BIG DATA, PRIVACIDAD Y PROTECCIÓN DE DATOS MEDIANTE LA ISO/IEC 27007:2017- APLICADO A LOS DATOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DEL NORTE. *Universidad Tecnica del Norte*, 29.
- Hernández Valadez, E. (2006). Algoritmo de clúster ing basado en entropía para descubrir grupos . *CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITECNICO NACIONAL*, <https://www.cs.cinvestav.mx/tesisgraduados/2006/resumenEdnaHernandez.html>.
- IBM. (17 de Agosto de 2021). *Conceptos básicos de ayuda de CRISP-DM*. Obtenido de IBM Documentacion: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- Imbaquingo, D., & Díaz, J. (2023). Quality and security as key factors in the development of Computer Audits in Higher Education Institutions. *Journal of Technology and Science Education. In Review*.
- INEC. (2021). *INEC*. Obtenido de Programa Nacion de Estadísticas 2017- 2021: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Normativas%20Estadísticas/Planificacion%20Estadística/Programa_Nacional_de_Estadística-2017.pdf
- ISO 25000. (2022). *ISO 25012*. Obtenido de Ecosistema para la Evaluación y Certificación de la Calidad del Producto Software: <https://iso25000.com/index.php/normas-iso-25000/iso-25012>
- Jones, S., Ross, S., & Ruusalepp, R. (2009). Data Audit Framework Methodology. *University of Glasgow*, 70.

- Kale, K. A. (2016). Review on Mining Association Rule from Semantic Data. *International Journal of Computer Science and Information Technologies*, Vol. 7, 1328-1331.
- Kamalpreet , B., & Anuranjan , M. (2017). A detailed study of clustering algorithms. *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 371-376.
- Kleboth, J. A., Kosorus, H., Rechberger, T., & Luning, P. A. (2022). Using data mining as a tool for anomaly detection in food safety audit data. *Food Control*, Volume 138, <https://doi.org/10.1016/j.foodcont.2022.109004>.
- Lara Torralbo, J. A. (2014). *Minería de Datos*. Madrid: Centro Estudios Financieros.
- Lemus Delgado, D., & Pérez Navarro, R. (2019). Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos. *Bogotá Universidad de los Andes*, 41-62.
- Liyanage, N. H. (2017). Advanced query model design concept to support multi-dimensional data analytics for relational database management systems. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 432-435.
- Mancero Menoscal, T. B. (2020). Detección de patrones de contrabando para la gestión de aprehensiones y retenciones, utilizando técnicas predictivas de clasificación y regresión de minería de datos. *Universidad Tecnica del Norte*, <http://repositorio.utn.edu.ec/handle/123456789/10860>.
- Manrique Plácido, J. M. (2019). *INTRODUCCIÓN A LA AUDITORÍA*. Chimbote: Universidad Católica Los Ángeles de Chimbote.
- Márquez Díaz, J. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. *Rev. Bioética y Derecho*, n.50, 315-331.
- Merilampi, S., Ihanakangas, V., & Virkki, J. (2019). User-driven development with scientific & applied research - RFID-controlled physiogame case study. *2019 IEEE International Conference on RFID Technology and Applications (RFID-TA)*.
- Molina López, J. M., & García Herrero , J. (2006). Técnicas de análisis de datos aplicaciones prácticas utiizando microsoft excel y weka. *Universidad Carlos III de Madrid*, 266.
- Mount , J., & Zumel, N. (2016). *Exploring Data Science*. Nueva York: Manning Publications.
- Moutaz Haddara, K. L. (2018). Applications of Big Data Analytics in Financial Auditing- A Study on The Big Four. *ACCOUNTING INFORMATION SYSTEMS (SIGASYS)*.
- Mustafa Ali, S., Anjum, N., & Kamel, M. (2018). Measuring management's perspective of data quality in Pakistan's Tuberculosis control programme: a test-based approach to identify data quality dimensions. *BMC Res Notes*, 11- 40.
- Naciones Unidas. (2021). *Objetivos de Desarrollo Sostenible*. Obtenido de Objetivos de Desarrollo Sostenible: <https://www.un.org/sustainabledevelopment/es/infrastructure/>
- Njiru, N. M., & Opiyo, E. (2018). Clustering and visualizing the status of child health in kenya: a data mining approach. *International Journal of Social Science and Technology*, <https://hal.science/hal-02265073>.
- Núñez Arias, E. B., González Nuñez, B. M., Nonell Fernández, L., & Rodríguez Pupo, J. M. (2021). CRISP-DM y K-means neutrosfía en el análisis de factores de riesgo de pérdida de audición en niños. *Revista Asociación Latinoamericana De Ciencias Neutrosóficas*, 73-81, <https://fs.unm.edu/NCML2/index.php/112/article/view/151>.

- Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe. (2021). *Plan nacional de desarrollo 2017-2021 toda una vida de Ecuador*. Obtenido de Plan nacional de desarrollo 2017-2021 toda una vida de Ecuador: <https://observatorioplanificacion.cepal.org/es/planes/plan-nacional-de-desarrollo-2017-2021-toda-una-vida-de-ecuador>
- Pacheco Curi, O. E. (2018). AUDITORIA DE SISTEMAS. En O. E. Pacheco Curi, *AUDITORIA DE SISTEMAS* (pág. 12). Lima: OEPC.
- Paetz, F., Steiner, W. J., & Hruschka, H. (2022). Advanced data analysis techniques with marketing applications. *Journal of Business Economics*, pages557–561.
- Paredes Murcia, A., & León Cárdenas, M. J. (2021). LA AUDITORÍA: FUENTE DE INFORMACIÓN ESTRATÉGICA EN LA INDUSTRIA HOTELERA1. *Turismo y Sociedad*, 207.
- Pedrero, V., Reynaldos Grandón, K., Ureta Achurra, J., & Cortez Pinto, E. (2021). Generalidades del Machine Learning y su aplicación en la gestión sanitaria en Servicios de Urgencia. *Revista médica de Chile*, 149(2), 248-254.
- Pérez López, C., & Santin González, D. (2007). *Minería de datos. Técnicas y herramientas*. Madrid: Ediciones Paraninfo, S.A.
- Pérez, J., Iturbide, E., & Olivares, V. (2015). A Data Preparation Methodology in Data Mining Applied to Mortality Population Databases. *Journal of Medical Systems volume*, 39, <https://doi.org/10.1007/s10916-015-0312-5>.
- Planas, M., Rodríguez, T., & Lecha, M. (2022). La importancia de los datos. *Nutrición Hospitalaria*, 11-13 http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-16112004000100003&lng=es&tlng=es.
- Plumed, F. M., Contreras Ochando, L., Ferri, C., & Hernández Orallo, J. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 3048-3061, doi:10.1109/TKDE.2019.2962680.
- Ponce Rodriguez, N. I. (2016). Auditoria a la gestion informatica con una vision de negocios integral en una empresa comercial. *Escuela Superior Politécnica del Litoral*, 108.
- Real Academia Española . (2023). *Procesamiento*. Obtenido de Real Academia Española : <https://dle.rae.es/procesamiento>
- Real Academia Española. (8 de Mayo de 2023). *definición 4*. Obtenido de estadístico: <https://dle.rae.es/estad%C3%ADstico>
- Redavid, D., Corizzo, R., & Malerba, D. (2018). An OWL Ontology for Supporting Semantic Services in Big Data Platforms. *2018 IEEE International Congress on Big Data (BigData Congress)*.
- Rendón Macías, M. E., Villasís Keeve, M. Á., & Miranda Novales, M. (2016). Estadística descriptiva. *Revista Alergia México*, 397-407.
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana*, 11-18.
- Rivas, G. A. (1988). Auditoría informática. Madrid: Diaz de Santos S.A.
- Robles Aranda, Y., & Sotolongo, A. R. (2013). Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL. *Journal of Information Systems and Technology Management*, 389-406.
- Ross, S. M. (2018). *Introducción a la estadística*. Barcelona: Editorial Reverte.
- Sánchez Ch, W. (2021). *Teoría de la auditoría*. Bogotá: Grupo Editorial Nueva Legislación SAS.

- Sanoran, b., & Ruangprapun, J. (2023). Initial Implementation of Data Analytics and Audit Process Management. *Sustainability*, <https://doi.org/10.3390/su15031766>.
- Santamaria Ruiz, W. (2006). Técnicas de minería de datos aplicadas en la detección de fraude: estado del arte. *Research Gate*, 1-8.
- Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, 2 - 420.
- Sdg Group. (16 de Noviembre de 2020). *Que es la analitica avanzada*. Obtenido de SdgGroup: <https://www.sdgggroup.com/es-ES/insights-room/que-es-la-analitica-avanzada>
- Sha, X. W., & Carotti Sha, G. (2016). Big Data. *AI & SOCIETY*, 1-4.
- Simić, S., Banković, Z., Villar, J. R., & Simić, D. (2020). A hybrid fuzzy clustering approach for diagnosing primary headache disorder. *Logic Journal of the IGPL*, vol. 29, 220-235.
- Sun, W., Cai, Z., Li, Y., & Liu, F. (2018). Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *Journal of Healthcare Engineering*, vol. 2018, 1 - 9 <https://doi.org/10.1155/2018/4302425>.
- Tamavo Alzate, A. (2001). Auditoría de Sistemas una Visión Práctica. *Universidad Nacional de Colombia*, 227.
- Tang, Y. X., Tang, Y. B., & Peng, Y. (2020). Automated abnormality classification of chest radiographs using deep convolutional neural networks. *npj Digital Medicine volume 3*, <https://doi.org/10.1038/s41746-020-0273-z>.
- Tejada Escobar, F., Murrieta Marcillo, R., & Villao-Santos, F. (2018). Big Data en la Educación: Beneficios e Impacto de la Analítica de Datos. *Revista Científica y Tecnológica UPSE*, 80-8.
- The University of Texas at Austin. (2023). *Audit Process*. Obtenido de Office of Internal Audits: <https://audit.utexas.edu/audit-process#reporting>
- Tsai, C. W., Lai, C. F., & Chao, H. C. (2015). Big data analytics: a survey. *Journal of Big Data*, 2-21.
- UNIR Ingeniería y tecnología. (2020). Análisis predictivo: ¿en qué consiste y para qué se usa? *UNIR REVISTA*.
- Veiga, N., Otero, L., & Torres, J. (2020). Reflexiones sobre el uso de la estadística inferencial en investigación didáctica. *InterCambios. Dilemas y transiciones de la Educación Superior*, 7(2), 94-106.
- Villalobos Murillo, J. (2006). AUDITANDO EN LAS BASES DE DATOS. *Universidad Nacional de Costa Rica*, 1-6.
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <http://www.jstor.org/stable/4132319>.
- Yang, H., Guohua, W., Wanlong, G., Qinghai, H., & Yixian, Y. (2022). Design and implementation of university audit platform based on big data analysis. *Procedia Computer Science, Volume 202*, <https://doi.org/10.1016/j.procs.2022.04.016>.
- Zambrano Montesdeoca, L. D. (2020). LA AUDITORÍA Y SU INFLUENCIA EN EL CONTROL DE LAS ACTIVIDADES REALIZADAS POR LAS EMPRESAS. *Universidad Técnica de Manabí*, 55.
- Zhang, L. (2021). Research on Audit Data Analysis under the Background of Big Data. *Journal of Electronic Research and Application*, Vol 5, No.3.
- Zhou, H., Wang, X., & Zhu, R. (2022). Feature selection based on mutual information with correlation coefficient. *Applied Intelligence*, 5457–547; <https://doi.org/10.1007/s10489-021-02524-x>.

Zhou, J. (2022). Application of Machine Learning Algorithms in Audit Data Analysis. *ICASIT 2021: 2021 International Conference on Aviation Safety and Information Technology*, 54–58.