



Modelo de pronósticos para indicadores operativos de auditoria en entidad de control basado en el uso de las técnicas de machine learning.

García Llanos, Carmen Isabel

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de Titulación previo, a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

Cárdenas Delgado, Sonia Elizabeth. Ph.D.

31 de julio de 2023



T-CARMEN GARCIA Julio_2023_mod.pdf

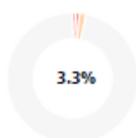
Scan details

Scan time: October 19th, 2023 at 3:47 UTC

Total Pages: 70

Total Words: 17399

Plagiarism Detection



Types of plagiarism		Words
Identical	0.7%	126
Minor Changes	0.8%	145
Paraphrased	1.4%	250
Omitted Words	10.5%	1827

AI Content Detection



Text coverage

- AI text
- Human text

🔍 Plagiarism Results: (22)

🌐 Garcia-Espana - Diseno e implementacion de un mo... 0.4%

<https://riunet.upv.es/bitstream/handle/10251/196062/garcia-...>

PABLO GARCÍA-ESPAÑA SIMÓ

UNIVERSITAT POLITÈCNICA DE VALÈNCIA Facultat de Administració y
Dirección de Empresas Diseño e implementación de un modelo de...

🌐 management construction sites - Traducción al esp... 0.4%

<https://www.linguee.es/ingles-espanol/traduccion/managem...>

Buscar en Linguee Sugerir como traducción de "management construction sites" Copiar ...

🌐 procurement and construction management - Trad... 0.4%

<https://www.linguee.es/ingles-espanol/traduccion/procurem...>

Buscar en Linguee Sugerir como traducción de "procurement and construction management" Copiar ...

Certified by

About this report
help.copleaks.com

copleaks.com

Ing. Sonia Cárdenas Delgado, Ph.D.

Director



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Certificación

Certifico que el trabajo de titulación “**Modelo de pronósticos para indicadores operativos de auditoria en entidad de control basado en el uso de las técnicas de machine learning**” fue realizado por la señora **García Llanos, Carmen Isabel**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 31 de julio de 2023

Ing. Sonia Cárdenas Delgado, Ph.D.

Director

C.C.:1713261160



**Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología
Centro de Posgrados**

Responsabilidad de Autoría

Yo **García Llanos, Carmen Isabel**, con cédula de ciudadanía n° 0201340882, declaro que el contenido de este trabajo de titulación: **“Modelo de pronósticos para indicadores operativos de auditoria en entidad de control basado en el uso de las técnicas de machine learning”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 31 de julio de 2023

García Llanos, Carmen Isabel

C.C.: 0201340882



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Autorización de Publicación

Yo **García Llanos, Carmen Isabel**, con cédula de identidad n° 0201340882, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Modelo de pronósticos para indicadores operativos de auditoria en entidad de control basado en el uso de las técnicas de machine learning”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 31 de julio de 2023

Garcia Llanos, Carmen Isabel

C.C.: 0201340882

Dedicatoria

Este trabajo está dedicado a mis hijos, Isabella y Alejandro, que son el motor en mi vida y son mi motivación para no rendirme en esta travesía y en la culminación de esta meta tan anhelada.

Agradecimiento

Agradezco a Dios por darme la vida y darme la fuerza para continuar en este proceso hasta llegar a la meta.

A mis padres Gerardo e Isabel, quienes desde el cielo han estado alentándome a seguir adelante y con su sabiduría me formaron, ellos que durante toda su vida me han dado el mejor ejemplo para ser la persona que soy.

A mi Tutora, Ing. Sonia Cárdenas Delgado, Ph.D., quien con su experiencia supo guiarme, fue de gran apoyo en todo momento y con su entereza me ayudó a alcanzar mi objetivo.

Mi agradecimiento a mi familia, mis profesores, mis amigos que de una u otra manera me brindaron su colaboración y apoyo.

Índice de Contenidos

Dedicatoria	6
Agradecimiento	7
Índice de Contenidos.....	8
Índice de Tablas	12
Índice de Figuras.....	13
Resumen.....	16
Capítulo I.....	18
Introducción.....	18
Antecedentes	18
Justificación, importancia y alcance.....	20
Planteamiento del Problema.....	21
Objetivo General.....	22
Objetivo Específicos	22
Preguntas de la Investigación.....	23
Hipótesis de Investigación	24
Estado del Arte	24
Criterios de inclusión y exclusión.....	25
Capítulo II.....	34
Marco Teórico	34
Verdaderos positivos (VP):	39

Verdaderos negativos (VN):.....	39
Falsos negativos (FN):.....	39
Falsos positivos (FP):	39
Metodología de la Investigación	44
Herramientas de Minería de Datos más usadas	46
RapidMiner	46
Weka	47
Orange	48
Knime	48
Sas	49
Comparación de herraminetas de data mining.....	49
Software utilizado	51
Hardware utilizado	52
Capítulo III.....	54
Desarrollo.....	54
Recopilación de la Información.....	55
Clasificación de la información en archivo físico	56
Tipos de Unidades y sus Ámbitos.....	58
Dna1 Dirección Nacional de Auditoria de Administración Central	58
Dna2 Dirección Nacional de Auditoria de Sectores Sociales	58
Dna3 Dirección Nacional de Auditoria de Deuda Pública y Finanzas.....	59

Dna4 Dirección Nacional de Auditoria de Telecomunicaciones, conectividad y sectores productivos	59
Dna5 Dirección Nacional de Auditoria de Gobiernos Autónomos Descentralizados	59
Dna6 Dirección Nacional de Auditoria de Recursos Naturales.....	60
Dna7 Dirección Nacional de Auditoria de Salud y Seguridad Social	60
Dna8 Dirección Nacional de Auditoria de Transporte, Vialidad, Infraestructura Portuaria y Aeroportuaria	60
Tipos de Informes	60
Indicadores de Control.....	61
Clasificación de la información en archivo digital	61
Selección de conjuntos de datos	66
Clasificación y Selección de Datos para estudio	68
Unidades y Ámbitos:.....	70
Tipos de exámenes:	70
Preprocesamiento de datos	71
Transformación.....	72
Minería de datos	73
Análisis de variables con Weka	96
Capítulo IV	106
Análisis de Resultados	106
Interpretación y Evaluación.....	106

Implantación y Despliegue.....	109
Capítulo V	111
Conclusiones.....	111
Recomendaciones.....	112
Bibliografía	112

Índice de Tablas

Tabla 1 <i>Listado de Estudios seleccionados por Grupo de Control</i>	26
Tabla 2 <i>Reconstrucción de la Cadena de Búsqueda</i>	30
Tabla 3 <i>Matriz de Confusión</i>	39
Tabla 4 <i>Comparación de Herramientas de Software para Minería de Datos</i>	49
Tabla 5 <i>Tipos de Examen de Auditoría</i>	56
Tabla 6 <i>Campos Usados para el Estudio</i>	67
Tabla 7 <i>Comparación de Resultados según las Matrices de Confusión generadas en los Modelos de Predicción</i>	95
Tabla 8 <i>Comparación de Resultados según las Matrices de Confusión generadas en los Modelos de Clasificación con Weka con otros campos y con otro banco de datos</i>	101

Índice de Figuras

Figura 1 <i>Representación de Variables Independiente y Dependiente</i>	34
Figura 2 <i>Representación del Descubrimiento de Conocimiento en Base de Datos. (Oded & Lior, 2010)</i>	44
Figura 3 <i>Mapa de Procesos seleccionados para el estudio</i>	54
Figura 4 <i>Consulta de datos del plan aprobado del año 2022</i>	62
Figura 5 <i>Consulta de datos del plan aprobado del año 2022</i>	62
Figura 6 <i>Consulta de la Ejecución de las acciones de control según el año 2022</i>	63
Figura 7 <i>Consulta de datos de Informes Aprobados del Año 2022</i>	63
Figura 8 <i>Consulta de los Informes Aprobados del Año 2022 con el Tiempo en días</i>	64
Figura 9 <i>Consulta de los Tiempos de Ejecución de la Orden de Trabajo, Tiempos de Aprobación del Informe</i>	65
Figura 10 <i>Datos seleccionados del Proceso de Planificación del Año 2022.</i>	66
Figura 11 <i>Datos seleccionados de la Ejecución del Control del año 2022</i>	67
Figura 12 <i>Informes Aprobados por Tipo de Examen</i>	71
Figura 13 <i>Tipos de Examen por Ámbito</i>	72
Figura 14 <i>Modelo de Clasificación de Datos con árboles de decisión en Knime</i>	73
Figura 15 <i>Nodo Excel Reader</i>	74
Figura 16 <i>Nodo String Replacer y su Configuración</i>	74
Figura 17 <i>Nodo Partitioning</i>	75
Figura 18 <i>Nodo Color Manager y Color Appender y su Configuración</i>	76
Figura 19 <i>Nodo Decision Tree Learner y su Configuración</i>	76
Figura 20 <i>Nodo Decision Tree Rulest y su Configuración</i>	77
Figura 21 <i>Nodo Decision Tree Predictor</i>	78
Figura 22 <i>Representación de la Configuración del Nodo Decision Tree Predictor</i>	78
Figura 23 <i>Nodo Scorer y su configuración</i>	79

Figura 24 <i>Resultado de la Matrix Confusión (Nodo Scorer) de Árboles de Decisión</i>	80
Figura 25 <i>Nodo Decision Tree Learner y su Configuración diferente</i>	80
Figura 26 <i>Resultado de la Matrix Confusión (Nodo Scorer) de Árboles de Decisión</i>	81
Figura 27 <i>Modelo de Naive Bayes</i>	82
Figura 28 <i>Nodo Excel Reader</i>	82
Figura 29 <i>Nodo String Replacer</i>	83
Figura 30 <i>Nodo Partitioning</i>	83
Figura 31 <i>Nodo Color Manager</i>	84
Figura 32 <i>Nodo Naive Bayes Learner</i>	84
Figura 33 <i>Configuraciones del Nodo Naive Bayes Learner</i>	84
Figura 34 <i>Nodo Naive Bayes Predictor</i>	85
Figura 35 <i>Configuraciones del Nodo Naive Bayes Predictor</i>	85
Figura 36 <i>Nodo Scorer</i>	86
Figura 37 <i>Configuración del Nodo Scorer</i>	87
Figura 38 <i>Resultado de la Matriz de Confusión (Nodo Scorer) de Naive Bayes</i>	87
Figura 39 <i>Modelo de Naive Bayes</i>	88
Figura 40 <i>Modelo K-Means</i>	89
Figura 41 <i>Nodo Excel Reader</i>	89
Figura 42 <i>Nodo Partitioning</i>	90
Figura 43 <i>Nodo K-Means</i>	90
Figura 44 <i>Configuración del Nodo K-Means</i>	90
Figura 45 <i>Nodo Color Manager</i>	91
Figura 46 <i>Nodo Shape Manager</i>	92
Figura 47 <i>Nodo Scatter Plot</i>	92
Figura 48 <i>Configuración del Nodo Scatter Plot</i>	92
Figura 49 <i>Resultado del nodo Scatter Plot de K-means</i>	93

Figura 50 <i>Modelo K-Means</i>	93
Figura 51 <i>Matriz de Confusión del Modelo de Arboles de Decisión</i>	94
Figura 52 <i>Matriz de Confusión del Modelo de Naive Bayes</i>	95
Figura 53 <i>Conjunto de Datos para el Análisis</i>	96
Figura 54 <i>Interface de la Herramienta Weka</i>	96
Figura 55 <i>Interface de carga de datos y análisis</i>	96
Figura 56 <i>Visualización con Unidad que ejecuta la Orden de Trabajo</i>	97
Figura 57 <i>Ejecución de la Orden de Trabajo según el ámbito</i>	97
Figura 58 <i>Ejecución de la Orden de Trabajo según la Provincia</i>	98
Figura 59 <i>Cantidad de Días de Ejecución de la Auditoría de la Orden de Trabajo</i>	98
Figura 60 <i>Cantidad de Días de Ejecutados de la Auditoría</i>	98
Figura 61 <i>Total de Días de Ejecución del proceso de la Auditoría</i>	99
Figura 62 <i>Matriz de Confusión del Modelo de Arboles de Decisión</i>	99
Figura 63 <i>Matriz de Confusión del Modelo de Naive Bayes</i>	99
Figura 64 <i>Matriz de Confusión del Modelo de JRip</i>	100
Figura 65 <i>Matriz de Confusión del Modelo de MultiClassClassifier</i>	100
Figura 66 <i>Modelo: Naive Bayes Classifier - Categorizado por Unidad</i>	101
Figura 67 <i>Modelo: Naive Bayes Classifier - Categorizado por Entidad</i>	102
Figura 68 <i>Modelo: Naive Bayes Classifier - Categorizado por Entidad</i>	103
Figura 69 <i>Modelo: MultiClass Classifier - Categorizado por Ámbito</i>	105
Figura 70 <i>Resultado del Gráfico de Dispersión (Nodo Scatter Plot) del Modelo K-means</i>	107

Resumen

En una entidad de control se busca medir el desempeño con base a una serie de indicadores operativos, entre ellos el tiempo promedio de realización de auditorías desde la planificación hasta el informe final, el cumplimiento de plazos, porcentaje de recomendaciones que se han implementado en las organizaciones. La tendencia actual y creciente en auditoría es la aplicación de técnicas de aprendizaje automático (machine learning) que busca mejorar la eficiencia, precisión y exhaustividad de los procesos de auditoría. Con estas técnicas se pueden analizar grandes volúmenes de datos para identificar patrones anómalos que podrían indicar errores o comportamientos inusuales.

Problema: La entidad no cuenta con un modelo que esté orientado al apoyo en la toma de decisiones informada, que aplique técnicas de aprendizaje automático y de auditoría, que permita identificar patrones de riesgo, anticipar problemas de incumplimiento en función de datos pasados, construir perfiles de comportamiento, clasificación de documentos relevantes reduciendo la carga de trabajo manual, y por supuesto evaluar la eficacia de la gestión. **Objetivo:** Implementar un modelo analítico-predictivo aplicando técnicas de machine learning para evaluar los niveles de cumplimiento mediante la categorización de los indicadores por tipo de examen y unidad establecidos por la entidad de control. **Metodología:** El presente trabajo aplica la metodología de Knowledge Discovery in Databases (KDD). **Resultados esperados:** El modelo analítico-predictivo permite construir modelos basados en indicadores para tener una visión a corto plazo de las operaciones de la entidad auditada y establecer controles de cumplimiento de los planes de trabajo en los equipos. Esto beneficia al proceso de toma de decisiones y la evaluación de la eficacia en la gestión.

Palabras clave: machine learning, audit, data mining, predictive algorithms

Abstract

In a control entity, the target is to measure the performance based on a series of operational indicators, including the average time to carry out audits from planning to the final report, the compliance with deadlines, and the percentage of feedbacks implemented in the organizations. The current and growing trend in auditing is the application of automatic learning techniques (machine learning) that seeks to improve the efficiency, precision, and wholeness of the audit process. According to these techniques, large volumes of data can be analyzed, to identify anomalous patterns. It could indicate errors or an unusual behavior. **Problem:** The entity does not have a model oriented to support informed decision-making, which allows automatic learning and auditing techniques to identify risk patterns, anticipating non-compliance problems based on past data, building behavior profiles, sorting relevant documents to reduce manual workload, and of course the assessment of an effective management. **Objective:** Implement an analytical-predictive model applying automatic learning techniques to verify compliance levels through categorized indicators by type of exam and units established in the control entity. **Methodology:** This research paper applies the Knowledge Discovery in Databases (KDD) methodology. **Expected results:** The analytical-predictive model allows the construction of models based on indicators to have a short-term vision for the operations of the audited entity and to establish compliance controls in support of the work plans in the teams. Therefore, the model benefits the decision-making process and the assessment of management effectiveness.

Key words: machine learning, audit, data mining, predictive algorithms

Capítulo I

Introducción

Antecedentes

Los indicadores son importantes para el seguimiento, la medición y la toma de decisiones informadas en diversos campos, como la gestión del control, auditoría, economía, el medio ambiente, la educación, entre otros. Los indicadores se utilizan para evaluar, medir y cuantificar diferentes aspectos de un proceso, fenómeno o sistema. La medición se puede realizar de manera cuantitativa o cualitativa.

Las actividades relacionadas con el proceso de auditoría en una organización, se cuantifican mediante indicadores de control que ayudan a identificar la calidad del trabajo realizado por el equipo de auditoría, según la eficacia y eficiencia de su desempeño. La selección de indicadores depende de los objetivos planteados, sin embargo, es necesario identificar los más relevantes y medibles para obtener la información más idónea para contribuir en la toma de decisión informada.

En una organización dedicada al control es necesario tener el control de todos sus procesos, para lo cual busca una serie de indicadores para cada uno de ellos. Para el presente trabajo de estudio que se refiere al proceso de control de auditoría, se consideran el periodo de tiempo programado versus el tiempo que conlleva la ejecución de una auditoría y contrastarlo con los plazos que rigen según la normativa vigente, se identifican patrones en auditorías periódicas, el cumplimiento de recomendaciones en las organizaciones que se han efectuado auditorías, entre otros. La elección de los indicadores depende del enfoque de la auditoría, de la organización, las áreas de enfoque en función de la organización que sea auditada.

Una tendencia actual y creciente en auditoría es la aplicación de técnicas de aprendizaje automático (machine learning) que busca mejorar la eficiencia, precisión y exhaustividad de los procesos de auditoría. Con estas técnicas se pueden analizar

grandes volúmenes de datos para identificar patrones anómalos que podrían indicar errores o comportamientos inusuales, lo cual es útil para detectar transacciones fraudulentas o actividades no autorizadas, fuera de la normativa o por omisión de alguna de ellas.

Los algoritmos de aprendizaje automático pueden clasificar transacciones en diferentes categorías, lo que puede ayudar a identificar patrones de riesgo o predecir riesgos futuros, anticipar problemas de cumplimiento o incumplimiento en función de datos pasados, detectar discrepancias sospechosas, construir perfiles de comportamiento, clasificación de documentos relevantes reduciendo la carga de trabajo manual, verificar cumplimiento de regulaciones, segmentación y categorización de las unidades de negocio, evaluar la eficacia de la gestión.

Es necesario tener conocimiento, comprensión y dominio profundo de las técnicas de aprendizaje automático y de auditoría para que se pueda aplicar como una herramienta de soporte a la toma de decisiones informadas; además, la calidad de los resultados depende de la calidad y cantidad de los datos utilizados para entrenar los modelos.

La entidad de control pretende aplicar las técnicas de aprendizaje automático y de auditoría, para disponer de una herramienta de soporte a la toma de decisiones informadas, con base a la información que se genera de los informes técnicos de los auditores, y que se encuentra en una base de datos.

La aplicación de las diferentes técnicas estimar el desempeño de los procesos y actividades relacionados con la función de control y auditoría en una organización. Se han determinado indicadores operativos para supervisar la eficacia, eficiencia y calidad del trabajo realizado por el equipo de auditoría. Estos indicadores permitirán medir el desempeño de los auditores, el tiempo promedio de realización de auditorías desde la planificación hasta el informe final, cumplimiento de plazos de acuerdo a lo programado,

porcentaje de órdenes de trabajo realizadas por unidad de control y ámbito, y tipos de informes.

Justificación, importancia y alcance

El tema desarrollado se basa en la necesidad que tiene la entidad de control de evaluar y medir el desempeño de los procesos y actividades relacionados con la función de auditoría en las diferentes organizaciones que son auditadas. Se propone desarrollar un modelo que, basado en el análisis de los datos, contribuya en la toma de decisiones informadas en la gestión del control y auditoría. Se determinan indicadores para evaluar, medir y cuantificar diferentes aspectos del proceso. La medición se puede realizar de manera cuantitativa o cualitativa.

El alcance de la presente investigación es realizar un estudio de las técnicas de análisis de minería de datos con el fin de encontrar el mejor modelo para contribuir en la toma de decisiones informadas en base a los indicadores de control. Aplicar técnicas de minería de datos con base a la información recopilada y analizada. Clasificar la información y seleccionar indicadores para el análisis previo de la investigación. En la entidad de control se busca medir el desempeño con base a una serie de indicadores operativos, entre ellos el tiempo promedio de realización de auditorías desde la planificación hasta el informe final, cumplimiento de plazos de acuerdo a lo programado, hallazgos que se repiten en auditorías sucesivas, porcentaje de recomendaciones de auditoría que se han implementado en las organizaciones. La elección de los indicadores depende del enfoque de la auditoría, de la organización, y las áreas de enfoque en función de la organización que sea auditada, de manera que se desarrolle un modelo dinámico.

Las técnicas de aprendizaje automático requieren ser estudiadas, aplicadas y probadas hasta encontrar el mejor modelo para una toma de decisiones informada que permita establecer cambios proactivos antes de que las condiciones hayan cambiado, de

esta manera la institución contará con una herramienta que le permita realizar mejoras en el proceso de manera proactiva y no solo reactiva.

Es importante considerar un análisis de este tipo en instituciones de control, en donde la aplicación de técnicas de análisis de datos y machine learning está iniciando, la presente propuesta puede ser considerada como un prototipo modelo de análisis predictivo que considera factores e indicadores propios de la necesidad institucional, que pueden servir de base para el análisis de otros procesos que tiene la institución, y que deben continuar en desarrollo. Por temas de seguridad y confidencialidad de la información que maneja la institución beneficiaria no se contempla describir todos los indicadores y el detalle de los resultados.

Planteamiento del Problema

Bajo la competencia de la entidad de control está la prevención y el buen uso de los recursos públicos, por lo que requiere implementar herramientas tecnológicas de vanguardia que se puedan aplicar a diferentes sectores. Para esto debe tomar las mejores decisiones en cuanto al control de sus procesos, por ejemplo, la ejecución de auditorías a las diferentes organizaciones públicas y el seguimiento al control del cumplimiento de las recomendaciones como resultado de cada examen, basados en toma de decisiones informadas; por ello la necesidad de identificar el número de indicadores necesarios.

Durante el proceso de auditoria se almacena gran cantidad de información en una base de datos institucional. Esta información demanda un análisis detallado en campos y contenidos registrados que están almacenados en varias bases de datos transaccionales. La cantidad de información es registrada en la base de datos institucional. Estos datos requieren ser analizados para detectar patrones que permitan realizar el control de cumplimiento de las recomendaciones para contribuir en la toma de decisiones de la alta gerencia.

Se requiere identificar los indicadores base para realizar un análisis adecuado, y tratamiento de los datos, aplicar técnicas de minería de datos para determinar patrones de

datos, predecir un comportamiento no adecuado, un posible fraude, entre otros, que se producen por desconocimiento de la Ley o por desconocimiento de los procedimientos o posibles actos de corrupción.

La entidad no cuenta con un modelo que esté orientado al apoyo en la toma de decisiones, que aplique técnicas de aprendizaje automático para mejorar la eficiencia, precisión y exhaustividad de los procesos de auditoría. Estas técnicas permiten analizar volúmenes de datos importantes para identificar patrones anómalos que podrían indicar errores o comportamientos inusuales, lo cual es útil para detectar actividades no autorizadas, fuera de la normativa o por omisión de alguna de ellas. En el modelo propuesto se considera la clasificación de las transacciones en diferentes categorías, lo que puede ayudar a identificar patrones de riesgo o predecir riesgos futuros, anticipar problemas de cumplimiento o incumplimiento en función de datos pasados, detectar discrepancias sospechosas, construir perfiles de comportamiento, clasificación de documentos relevantes reduciendo la carga de trabajo manual, verificar cumplimiento de regulaciones, segmentación y categorización de las instituciones auditadas, y por supuesto evaluar la eficacia de la gestión.

Objetivo General

Implementar un modelo analítico-predictivo aplicando técnicas de machine learning para evaluar los niveles de cumplimiento mediante la categorización de los indicadores por tipo de examen y unidad establecidos por la entidad de control.

Objetivo Específicos

OE1: Ejecutar una revisión de literatura para establecer las técnicas de minería de datos más apropiadas para el organismo de control.

OE2: Seleccionar los indicadores sensibles en la operación auditada de la entidad.

OE3: Realizar una validación de datos, verificando consistencia y calidad, previo a ser utilizado en el modelo desarrollado de minería de datos.

OE4: Implementar un modelo analítico-predictivo mediante técnicas de ML para construir modelos predictivos basados en indicadores de auditoría, que permita tener una visión a corto plazo de la operación de la entidad auditada.

OE5: Validar el modelo predictivo, a través de los resultados obtenidos, de patrones y tendencias de los indicadores del modelo propuesto, y compararlos con resultados obtenidos de forma manual, para determinar el nivel de confianza del modelo propuesto.

Preguntas de la Investigación

En el presente estudio, se plantea la implementación de técnicas de minería de datos, identificando indicadores de control con el propósito de aportar a la toma de decisiones informadas, para lo cual se planteó las siguientes preguntas de estudio por cada uno de los objetivos:

OE1 – RQ1.1: ¿Qué tipos de estudios relacionados con auditoría aportan en la definición de las mejores técnicas de minería de datos?

OE1 – RQ1.2: ¿Cuáles son las características que tienen las técnicas de minería de datos implementadas en el ámbito de la auditoría y control?

OE2 – RQ2.1: ¿Cuáles son los indicadores relacionados con la normativa legal vigente que se aplican en la ejecución del de la operación del control?

OE2 - RQ2.2: ¿Qué técnicas de los valores atípicos mejor se adaptan a la realidad de la institución y que permiten asegurar la calidad de la información?

OE3 – RQ3.1: ¿Cuál es la técnica de minería de datos que mejor se adapta a la institución y cuáles son las características que hicieron seleccionar esta técnica?

OE3 – RQ3.2: ¿Es posible implementar un modelo de minería de datos con la técnica seleccionada para los datos del proceso seleccionado de la Entidad de Control?

OE4 – RQ4.1: ¿Cuál es el periodo de tiempo a ser aplicado el modelo de minería de datos, según los indicadores de auditoría y control, con mejores resultados a corto plazo?

OE4 – RQ4.2: ¿Cuál es el porcentaje de error que se debe aceptar al implementar modelos de minería de datos?

OE4 – RQ4.2: ¿Aplicando las técnicas de minería de datos, es posible validar el modelo propuesto?

OE5 – RQ5.1: ¿Es posible determinar un nivel de confianza para los resultados obtenidos, aplicando el modelo propuesto, y que permita demostrar una mejora en los mismos?

Hipótesis de Investigación

La aplicación de las técnicas de Minería de Datos, permitirán tener los pronósticos en los indicadores de auditoría y control de la Institución.

Señalamiento de Variables:

Variable dependiente: Predicciones en los Indicadores de auditoría

Variable independiente: Técnicas de aprendizaje supervisado

Con la hipótesis planteada, se consideran actividades que permiten medir el desempeño y determinar pronósticos a los resultados.

Se plantea una hipótesis que considera elementos de naturaleza descriptiva que son cuantificables lo que permitirá establecer predicciones en los resultados.

Para demostrar la hipótesis planteada, se utilizó varias consultas sql que fueron elaboradas, afinadas y ejecutadas, para posteriormente ejecutar y evaluar los resultados del modelo analítico-predictivo aplicando técnicas de validación de minería de datos, como por ejemplo la validación cruzada (Microsoft, 2022), matriz de confusión (Microsoft, 2022), y poder determinar la técnica más precisa para encontrar el nivel de confianza esperado.

Estado del Arte

Para el estado del arte, se realizó un análisis sistemático de la literatura existente valiéndose de los criterios de inclusión, exclusión y la estrategia de búsqueda, la fuente de

búsqueda de la información utilizada para la presente investigación fue el repositorio Google Académico.

Criterios de inclusión y exclusión

Para determinar los criterios de exclusión e inclusión, se toman en cuenta los estudios con las características mas cercanas al tema propuesto.

Criterios de inclusión

- Se seleccionaron artículos científicos que tengan relación con trabajos de investigación para la toma de decisiones
- Que los artículos seleccionados contengan información relacionada con temas de auditoría y el uso de técnicas de inteligencia de negocios.
- Que el articulo seleccionado contenga información relacionada con la aplicación de técnicas de minería de datos.
- Que el artículo seleccionado tenga relación con la implementación de herramientas de minería de datos.
- Se incluyen trabajos de investigación a partir del año 2012.

Criterios de exclusión

- Se excluyen artículos que no estén relacionados con el tema de investigación.
- Los trabajos de investigación anteriores al año 2012.
- Se excluye los artículos en idiomas diferentes al inglés o español.

Definición de la estrategia de búsqueda

Para la selección y definición de la estrategia de búsqueda se realizó lo siguiente:

En primer lugar, se realizó la ejecución de la **revisión inicial de literatura**, con la validación de estudios relacionados al tema planteado.

En segundo lugar, se ejecuta **la validación cruzada**, lo cual garantiza que los

estudios cumplan con los criterios de exclusión e inclusión, y conforman lo que se conoce como el listado de integración del grupo de control.

Finalmente, en tercer lugar, tenemos el listado de los estudios que conforman la ***Integración del Grupo de Control***, para conseguir este listado, se realizó una validación y análisis inicial del título de los estudios, introducción, conclusiones y palabras claves, se puede indicar que estos estudios cumplen con las características de la investigación.

Los estudios que conforman el grupo de control se detallan en la Tabla 1.

Tabla 1

Listado de Estudios seleccionados por Grupo de Control.

Grupo de Control	Título	Palabras clave
EstC1	Deep Learning and the Future of Auditing: How an Evolving Technology Could transform Analysis and Improve Judgment.	Artificial Intelligence; Deep Learning; Audit; Evolving Technology; Transform Analysis
EstC2	Auditor Choice Prediction Model Using Corporate Governance and Ownership Attributes: Machine Learning Approach	Auditor Choice, Machine Learning, Prediction. Comparison study; Gaussian process regression;
EstC3	An Empirical Comparison of Machine Learning Models for Time Series Forecasting	Machine learning models; Neural network forecasting; Support vector regression
EstC4	Using Deep Learning Algorithms for CPAs' Going Concern Prediction.	Going concern prediction; deep learning; deep neural networks (DNN); recurrent neural network (RNN); classification and regression tree (CART)
EstC5	Continuous auditing and data mining for strategic risk control and anticorruption: Creating "fair" value in the digital age	Data Mining Methods, audit risk, risk management, audit institutions, reports
EstC6	Financial reporting fraud detection: An analysis of data mining algorithms	Financial reporting fraud; fraud detection; fraud predictors; classification algorithms

Los estudios revisados se detallan a continuación:

[EstC1] (Sun & Vasarhelyi, Deep Learning and the Future of Auditing: How an Evolving Technology Could Transform Analysis and Improve Judgment., 2017) The Future of Auditing and Deep Learning: How an Evolving Technology Could Transform Analysis and Improve Judgment.

Este artículo nos presenta la tecnología de deep learning o de aprendizaje profundo, como una forma emergente de inteligencia artificial que se puede entrenar para descubrir patrones en grandes volúmenes de datos lo que serían imposible de procesar para los cerebros humanos. Esta tecnología está aún en evolución presenta una manera de utilizar big data para crear evidencia de auditoría complementaria con el fin de mejorar la eficacia y la eficiencia de la automatización de auditoría y la toma de decisiones.

[EstC2] (Rahman, Masrom, Zakaria, & Halid, 2021) Prediction Model of Auditor Choice Using Corporate Governance and Ownership Attributes: Machine Learning Approach

Este artículo nos presenta los conocimientos fundamentales para la implementación de un modelo de aprendizaje automático, donde el auditor escoge los procesos de los informes financieros y la calidad de la auditoría. Este estudio se basa en cuatro algoritmos seleccionados que se prueban en un conjunto de datos real de 2.262 observaciones que cotizaron en la bolsa de valores de Malasia entre los años 2000 y el 2007. El funcionamiento de cada uno de los algoritmos de aprendizaje automático en el conjunto de datos de seleccionado es observado en las tres categorías seleccionadas, las características de la empresa, la gobernanza y la propiedad. Los hallazgos muestran que los modelos de aprendizaje automático presentan un mejor rendimiento de precisión con la categoría de las características de propiedad en especial con el modelo de Naïve Bayes.

[EstC3] (Nesreen, Amir, Neamat, & Hisham, 2010) Empirical Comparison of Machine Learning Models for Time Forecasting Series.

El artículo indica que se realizó un estudio de comparación de los principales modelos de aprendizaje automático con un gran volumen de datos, entre los modelos usados están, las series de tiempo, redes neuronales Bayesianas, redes neuronales de regresión, los modelos multicapa, funciones de base radial, regresión del K-vecino más cercano, árboles de regresión CART, regresión de vectores de soporte y procesos de Gauss. Este estudio revela diferencias significativas entre los diferentes métodos, los mejores modelos resultaron ser dos, el de regresión del proceso gaussiano y el modelo multicapa. Finalmente, el estudio nos muestra que además de la comparación entre modelos, se ha probado diferentes métodos de preprocesamiento y se ha demostrado que tienen diferentes impactos en el rendimiento.

[EstC4] (Jan, 2021) Algorithms Deep Learning for CPAs'.

Este artículo nos indica que las opiniones de empresas de auditoría a las cuentas públicas certificadas, son la base para el control de las organizaciones para poder conocer si éstas realizan operaciones dentro de la normativa legal y sostenibles en el tiempo. El objetivo del estudio es la construcción de modelos predictivos para ayudar a los auditores y contadores públicos puedan dar juicios mas acertados respecto a los negocios en marcha, ejecutando algoritmos de aprendizaje profundo, se usaron 3 métodos, redes neuronales profundas, redes neuronales recurrentes y árbol de clasificación y regresión. Según los primeros resultados, con las variables importantes seleccionadas por los algoritmos antes mencionados tiene la mejor precisión de predicción, es así que la precisión del conjunto de datos de prueba es del 95,28 % y la precisión promedio es del 93,92 %.

[EstC5] (Muslihatun, Hantono, & Fauziati, 2020) Auditoria continua y minería de datos para control de riesgos estratégicos y anticorrupción.

Este artículo tiene como objetivo el estudio mediante la investigación de un caso de auditoría continua para anticorrupción. La evolución de la tecnología puede ofrecer valiosas

oportunidades para integrar controles de legalidad y procesos comerciales que sean consistentes con el creciente llamado a combatir la corrupción a nivel institucional, los investigadores han propuesto marcos conceptuales que demuestran las ventajas visibles de la auditoría continua y la minería de datos, pero aún existen dificultades significativas en la práctica. Los investigadores indican que se necesitan más estudios con la aplicación de más casos de estudio sobre este tema.

Para el tema específico de anticorrupción, en este estudio se demuestra el impacto práctico de la auditoría continua y la minería de datos para el control de riesgos estratégicos, según Chan y Vasarhelyi (2011). Los resultados muestran que la efectividad de la auditoría continua depende de las características seleccionadas más influyentes, como un enfoque integrador de gestión del cambio y de la visión estratégica, el mapeo de riesgos, manteniendo una cultura de no corrupción. El artículo nos dice que existe la necesidad de crear valor legal en la era digital.

[EstC6] (Mohammadi, Yazdani, Khanmohammadi, & Maham, 2020) Financial reporting fraud detection: An analysis of data mining algorithms

En la última década, se descubrieron y denunciaron fraudes financieros de alto perfil cometidos por grandes empresas tanto en países desarrollados como en desarrollo. Este estudio compara el rendimiento de cinco modelos estadísticos y de aprendizaje automático populares para detectar el fraude en los estados financieros. Los objetos de investigación son empresas que experimentaron estados financieros fraudulentos y no fraudulentos entre los años 2011 y 2016. Los algoritmos aplicados fueron, red neuronal artificial, red bayesiana, análisis discriminante, regresión logística y máquina de vector de soporte, los resultados de la aplicación de estos algoritmos muestran que la red neuronal artificial funciona mejor frente al resto de algoritmos aplicados.

Los resultados también revelan cierta diversidad en los predictores utilizados en los algoritmos de clasificación. De los 19 predictores examinados, solo nueve son seleccionados y

utilizados consistentemente por diferentes algoritmos de clasificación: productividad de los empleados, cuentas por cobrar a ventas, Deuda a capital, inventario a ventas, ventas a activos totales, retorno sobre capital, retorno sobre ventas, pasivos a gastos por intereses y activos a pasivos. Estos hallazgos amplían la investigación del fraude en los estados financieros y pueden ser utilizados por profesionales y reguladores para mejorar los modelos de riesgo de fraude.

Cadena de búsqueda

Para armar la cadena de búsqueda se analizó los artículos mencionados en el grupo de control, se buscan palabras que se repiten entre las investigaciones y se seleccionan las palabras

comunes entre los estudios y óptimas referentes al objetivo del presente estudio, para lo cual se definen los siguientes enfoques: Audit, data mining, predictive algorithms, machine learning que se pueden observar en la Tabla 2.

Tabla 2

Reconstrucción de la Cadena de Búsqueda.

Enfoque	Palabras clave	EstC 1	EstC 2	EstC 3	EstC 4	EstC 5	EstC 6	Número de repetición es
	Deep learning	X			X	X		3
	Corporate Governance		X					1
Machine learning	Machine learning models		X	X				2
	Time series			X				1
	Data mining						X	1

	Audit	X	X		X		3
Audit	Risk control				X		1
	Fraud detection					X	1
	Forecasting			X			1
	Predictive análisis	X					1
Predictive	Time series			X			1
	Predictive models		X		X	X	3

Según el análisis de los resultados de la combinación de los estudios de control que se muestran en la Tabla 2, se formó la cadena de búsqueda con la estructura de los textos de las palabras clave que más se repiten, para determinar la cadena de búsqueda se utilizó los conectores OR y AND, OR se usó para unir las palabras claves relacionadas y AND para concatenar estas uniones, resultando la siguiente cadena de búsqueda:

((deep learning OR machine learning) AND (audit OR audit indicators) AND predictive models)

Finalmente, se consideró otra condición para filtrar los estudios relacionados con la investigación planteada, se ejecutó esta cadena de búsqueda en Google Academic y aplicando los criterios de inclusión y exclusión se encontraron 6 estudios referentes a la problemática de investigación relacionando con el planteamiento del uso de técnicas de minería de datos como una de las posibles soluciones.

Se aplicó el estado del arte, que según Schwarz es considerado como el conocimiento más actualizado que existe para resolver un problema de estudio propuesto, el estado del arte considera los conocimientos e investigaciones más recientes que se acercan a una solución

relacionada al problema de investigación o han apoyado con algún aspecto o han guiado en la solución del mismo (Schwarz, 2013)

Se observa que el estado del arte muestra nuevo conocimiento, que permite verificar el nivel de actualización del tema planteado para la investigación, cual es su utilidad, esto nos ayuda a descubrir en la revisión de literatura, quien investigó sobre el tema, porque se ha realizado, para que se usó, donde y cuando se han hecho estudios relacionados con nuestro tema de investigación. Un estado del arte bien realizado, guía al investigador hacia la solución del problema de investigación al observar soluciones relacionadas con el problema de investigación planteado, que han logrado otros investigadores.

Según el estudio del arte revisado, se puede concluir, a nivel internacional hay propuestas para el uso de técnicas de minería de datos que tienen un nivel de madurez asertivo en el campo de la analítica predictiva; en cuanto al problema de investigación planteado se verificaron estudios en los que es imprescindible la automatización de los valores que son necesarios predecir además de considerar rangos históricos con consistencia de datos nos permite encontrar mejores resultados, los estudios revisados a nivel nacional enfocados al sector de auditoría de la Entidad de Control son escasos, estos estudios se enfocan más en los negocios, en el análisis financiero enfocado en el cliente, planteando posibles soluciones con el uso de técnicas de minería de datos. Así mismo otros estudios revisados usan análisis estadístico y se orientan en el área de riesgos tomando en cuenta un análisis de las variables de riesgo, se plantean sugerencias en el uso de herramientas de minería de datos para el análisis de indicadores.

No se encontraron estudios que se apliquen en un modelo de pronósticos para entidades de control ecuatorianas.

En consecuencia, el presente estudio contribuirá con la generación de conocimiento sobre la problemática planteada en el ámbito del proceso de auditoría de la Entidad de Control, a través de una perspectiva más integral desde el proceso de captura de

información, consolidación automática y propondrá un modelo de pronóstico basándose en el análisis y uso de técnicas de minería e datos aportando con nuevos lineamientos de análisis a los estudios actuales, tomando en cuenta las variables y características de la gestión de control ecuatoriana.

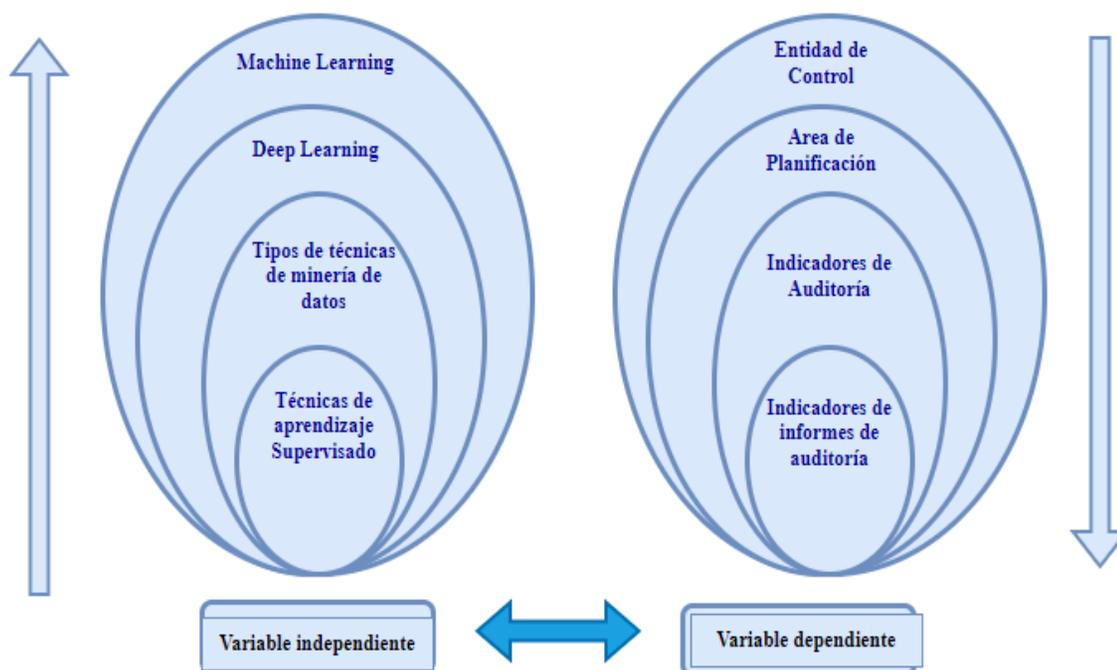
Capítulo II

Marco Teórico

En este capítulo se describe la fundamentación teórica que ha sido estudiada para utilizar como base para el desarrollo del trabajo planteado. Para explicar el análisis de la teoría, se lo realizó usando las variables identificadas del estudio, con el fin de investigar jerárquicamente cada uno de sus niveles hasta llegar al nivel que describe y explica las variables independientes y dependientes del tema de estudio, para lo cual se plantea una jerarquía de estudio que se representa en la Figura 1.

Figura 1

Representación de Variables Independiente y Dependiente



Fundamentación de las variables Independientes

Machine Learning

Esta es un método científico del ámbito de la Inteligencia Artificial que genera sistemas que aprenden automáticamente. Aprender en este contexto significa descubrir patrones complejos en grandes volúmenes de datos. El motor que revisa los datos tiene la capacidad de pronosticar los comportamientos futuros (Reyes, 2020).

Deep Learning

El Deep learning o aprendizaje profundo permite a los modelos computacionales que se componen de múltiples capas de procesamiento aprendan representaciones de datos con múltiples grados de abstracción. Estos métodos van mejorando drásticamente, entre ellos tenemos el reconocimiento de voz, el reconocimiento de objetos visuales, la detección de objetos y muchos otros dominios, como el descubrimiento de fármacos y la genómica, entre otros. (LeCun, Bengio, & Hinton, 2015).

Tipos de técnicas de Minería de Datos

La Minería de Datos, consiste en el proceso de descubrimiento de patrones o de conocimiento útil desde fuentes de datos que pueden ser bases de datos, textos, imágenes, la Web, archivos Excel, entre otras. Una aplicación de minería de datos por lo general inicia con un entendimiento y un análisis de datos, entendimiento y dominio de aplicación, se identifican las fuentes de donde provienen estos datos y los datos objetivos (Herrera, Ruiz, Romagnano, & Lund, 2017).

Entre las técnicas más comunes tenemos, las de aprendizaje supervisado o clasificación, aprendizaje no supervisado o clustering, reglas de asociación, y patrones secuenciales.

Técnicas Supervisadas

Para las técnicas supervisadas, existen modelos de aprendizaje supervisado que se aprenden funciones, relaciones que asocian entradas con salidas, por lo que se ajustan a un conjunto de ejemplos de los que conocemos la relación entre la entrada y la salida deseada.

Considerando el tipo de salida, suele definirse una subcategoría que diferencia entre modelos de clasificación y modelos de regresión, los primeros si la salida es un valor categórico, por ejemplo, una enumeración, o un conjunto finito de clases, y los segundos, si la salida es un valor de un espacio continuo (Sancho Caparrini, 2020).

Las técnicas supervisadas son de predicción y clasificación

Las técnicas supervisadas de predicción se consideran:

- Regresión
- Árboles de predicción
- Estimador de núcleos

Las técnicas supervisadas de clasificación se consideran:

- Tabla de decisión
- Árboles de decisión
- Inducción de reglas Bayesianas
- Basado en ejemplares
- Redes neuronales
- Lógica borrosa
- Técnicas genéticas

(Molina López & García Herrero, 2006)

Árboles de decisión, es una de las técnicas supervisada de las más eficaces para la clasificación, es de entendimiento e interpretación amigable, la presentación de resultados es de forma gráfica, utiliza la lógica booleana, este algoritmo es capaz de combinar métodos matemáticos con el uso de la computación, de esta manera logra tener una categorización,

aporta en la descripción y generalización del conjunto de datos. Con esta técnica se analizan grandes volúmenes de datos, para lo cual utiliza recursos informáticos básicos en un periodo razonable de tiempo, por otro lado el algoritmo está apto para manejar datos numéricos al igual que datos categorizados. (Montenegro Fierro, 2020).

Naive Bayes, es una de las técnicas supervisada de clasificación más usada, se basa en probabilidades condicionadas con datos conocidos, es de uso amigable y rápida. Este algoritmo se basa en calcular probabilidades de datos conocidos y la aplicación de una formula, es así como nos proporciona los resultados (Montenegro Fierro, 2020).

Técnicas no Supervisadas

Las técnicas de aprendizaje no supervisado son aquellas en las que se preocupan en aumentar el conocimiento estructural de los datos disponibles y otros posibles datos futuros que provengan del mismo fenómeno, es decir se buscan agrupaciones de datos según su semejanza, esto se conoce como la técnica de clustering, se acortan las estructuras de datos manteniendo sus características principales. Esta técnica no se interesa por ajustar pares como la entrada y salida. (Sancho Caparrini, 2020).

Las técnicas no supervisadas son de clustering y asociación

Entre las técnicas no supervisadas de clustering tenemos:

- Numérico
- Conceptual
- Probabilístico

Las técnicas no supervisadas de asociación se consideran:

- A priori

Se usan generalmente para la extracción de información útil a partir de grandes volúmenes de datos, usan los métodos descriptivos para buscar patrones interpretables para describir datos.

Las técnicas no supervisadas o descriptivas, se basan en los datos actuales para descubrir patrones y tendencias, no utilizan datos históricos. (Herrera, Ruiz, Romagnano, & Lund, 2017).

En las técnicas no supervisadas de clusterización está el algoritmo K-Means, éste agrupa determinadas características y divide los elementos en k categorías, ésta concentración va considerando las distancias entre cada elemento y el centroide de cada grupo o cluster, para lo cual usa la distancia cuadrática (Datos, 2021).

Dentro de los algoritmos de clasificación utilizados en minería de datos se pueden mencionar:

Las curvas ROC (característica operativa del receptor), que permite evaluar el rendimiento del modelo de machine learning de dos clases, mediante la representación gráfica de la relación entre el porcentaje de Verdaderos Positivos y el porcentaje de Falsos Positivos.

Se tiene como línea base a los valores indicados en la Matriz de Confusión, y a partir de ellos se puede contar con una serie de medidas útiles para considerar en nuestro análisis.

$$\text{Precisión} = (VP + VN) / (VP + FP + FN + VN)$$

$$\text{Sensibilidad} = VP / (VP + FN)$$

$$\text{Especificidad} = VN / (VN + FP)$$

Considerando la estadística, y tomando un estimador de una variable, el mismo que contempla un parámetro ajustable que le permite especificar sus curvas de Sensibilidad y Especificidad o Curvas ROC.

En la Curva ROC se grafica la sensibilidad del modelo frente al valor que se obtiene de restar la especificidad de la unidad, es decir 1-especificidad.

Se definen dos conceptos: Tasa de Verdaderos Positivos y Tasa de Falsos Positivos.

$$\text{Tasa de Verdaderos Positivos (TVP)} = \text{Sensibilidad} = VP / (VP + FN)$$

$$\text{Tasa de Falsos Positivos (TFP)} = FP / (FP + VN)$$

Como se señaló anteriormente, la sensibilidad se evalúa frente al restar a una unidad la especificidad, por lo que también aplica $TFP = 1 - \text{Sensibilidad} = 1 - VP / (VP + FN)$ (Melillanca, 2018).

Scorer o Matriz de confusión, función de Minería de clasificación, que permite visualizar la distribución de los registros en términos de clases reales y clases predichas.

A continuación, se presenta la Matriz de Confusión, en la Tabla 3:

Tabla 3

Matriz de Confusión

		Valor real	
		Positivos	Negativos
Valor predicho	Positivos	Verdaderos Positivos (VP)	Falsos Positivos (FP)
	Negativos	Falsos Negativos (FN)	Verdaderos Negativos (VN)

Los datos que figuran en la matriz de confusión, se describen a continuación:

Verdaderos positivos (VP): Pertenecen a los valores predichos como Positivos por el modelo (clase 1) de clasificación y que justamente corresponden a un valor Positivo, para los datos de prueba.

Verdaderos negativos (VN): Son los valores pronosticados como Negativos por el modelo (clase 0) y que pertenecen a un valor Negativo en el conjunto de datos.

Falsos negativos (FN): Pertenecen a valores pronosticados por el modelo como Negativos de forma incorrecta, ya que en el conjunto de datos corresponden a un valor Positivo.

Falsos positivos (FP): Son los valores pronosticados como Positivos por el modelo, pero que en los datos de prueba pertenecen a valores Negativos (Melillanca, 2018).

Lift chart(gráfico de elevación), la curva aleatoria siempre se representa como una línea horizontal, en el punto del diagrama en el que la curva del modelo desciende por debajo de la curva aleatoria, el factor de elevación es inferior a 1. Esto significa que los registros del modelo contienen menos valores de destino que un muestreo aleatorio con valores de destino distribuidos por igual.

Gráfico de dispersión o Scatter plot, ésta gráfica nos permite visualizar, los problemas o causas relacionados con la calidad de los datos de dos variables que se relacionan entre sí. Los valores que se toman del conjunto de datos seleccionado se representan como puntos de un plano cartesiano y se muestra la relación como positiva, cuando los valores crecen juntos, como negativa cuando un valor disminuye a medida que el otro aumenta, como nulo cuando un valor no tiene correlación lineal, como exponencial, cuando los valores van en forma de U, pudiendo ser la fuerza de correlación fuerte, débil o ninguna.

Bar chart o grafico de barras, este gráfico visualiza una o más métricas agregadas para diferentes particiones de datos con barras rectangulares donde las alturas son proporcionales a los valores de las métricas. Las particiones están definidas por los valores en una característica categórica. Este flujo de trabajo muestra diferentes formas de usar el nodo Gráfico de barras de JavaScript. Dados los datos que contienen una variable categórica, este nodo crea una gráfica que muestra diferentes estadísticas de atributos numéricos dentro de cada categoría. Las estadísticas disponibles son recuento de ocurrencias, suma y promedio.

Minería de texto, esta técnica hace análisis de textos, para lo cual se preocupa en convertir un texto no estructurado en datos estructurados. Esta técnica de minería toma la metodología del “procesamiento natural del lenguaje” o “**Natural Language Processing**”, que permite que las máquinas traten los datos como con el lenguaje humano y lo simulen de manera automática. (DataScientest, 2022).

Actualmente, la **minería de texto** puede aprovechar de algoritmos automatizados para procesar documentos digitales no estructurados de forma rápida y eficiente, se puede también desarrollar sus propios métodos con la finalidad de organizar los documentos digitales y poder analizarlos (Contreras Barrera, 2016).

Fundamentación de la variable dependiente

Entidad de Control

En la Constitución de República del Ecuador, artículo 211 de la ley, la Entidad de Control “es un organismo técnico encargado del control de la utilización de los recursos estatales, y la consecución de los objetivos de las instituciones del Estado y de las personas jurídicas de derecho privado que dispongan de recursos públicos”. (Asamblea Nacional del Ecuador, 2008, pág. 44)

En la Constitución de República del Ecuador, artículo 212 de la ley, Serán funciones de la Entidad de Control, además de las que determine la ley:

“1. Dirigir el sistema de control administrativo que se compone de auditoría interna, auditoría externa y del control interno de las entidades del sector público y de las entidades privadas que dispongan de recursos públicos.

2. Determinar responsabilidades administrativas y civiles culposas e indicios de responsabilidad penal, relacionadas con los aspectos y gestiones sujetas a su control.

3. Expedir la normativa para el cumplimiento de sus funciones.

4. Asesorar a los órganos y entidades del Estado cuando se le solicite.”

(Asamblea Nacional del Ecuador, 2008, pág. 44).

Área de Planificación

El Área de Planificación, es una de las unidades administrativas de la Entidad de Control que se encarga de la coordinación de la planificación y ejecución del control de la utilización de los recursos estatales, organizando el control hacia todos los sectores del Estado.

La Entidad de Control cuenta con Unidades de Auditoría para control externo e interno, las Unidades de Auditoría para Control Externo son las instauradas en la entidad de control que efectúan la fiscalización a las diferentes entidades del estado, así mismo, las Unidades de Auditoría para el Control Interno, son las que cuentan con una Unidad de Auditoría Interna en las diferentes entidades del estado y que efectúan el control a la entidad donde se encuentran, sin embargo dependen técnica y administrativamente de la Contraloría General del Estado.

Indicadores de auditoría

Los indicadores de auditoría son de gran importancia porque nos permiten identificar debilidades, fortalezas y deficiencias en la institución, su estado actual y real, que permite sacar conclusiones, planear y tomar decisiones en un momento indicado y nos permita evaluar los resultados, la gestión y la ejecución de las acciones de control realizados por la institución, emitir diagnósticos, establecer lineamientos para tener proyecciones y prever las acciones a ejecutar y que podrán ser visualizadas en los informes de auditoría, según que sucede en el transcurso del tiempo.

Para la Entidad de Control los indicadores son de suma importancia, ya que a través de ellos se mide los diferentes procesos que se ejecutan, por ejemplo, el análisis de auditoría, el estado de una acción de control, de informe de auditoría, y de otros procesos según los resultados de ejecución, mediante los indicadores permite conocer la eficiencia y así establecer y conocer la capacidad de ejecución de los diferentes procesos y en especial para este trabajo nos permite conocer de las acciones de control planificadas. En la Entidad de Control durante la investigación se pudo observar otros indicadores y o factores que afectan la gestión de la institución y que pueden ser considerados.

Indicadores de informes de auditoría

Para determinar los indicadores de los informes de auditoría, se revisó la Ley Orgánica de la Contraloría General de Estado ecuatoriano, en especial los artículos 18, 19 y 26 que menciona lo siguiente:

“Art. 18.- Alcance y Ejecución de la Auditoría Gubernamental.- El control externo que realizará la Contraloría General del Estado se ejercerá mediante la auditoría gubernamental y el examen especial, utilizando normas nacionales e internacionales y técnicas de auditoría.

La auditoría gubernamental realizada por la Contraloría General del Estado, consiste en un sistema integrado de asesoría, asistencia y prevención de riesgos que incluye el examen y Evaluación críticos de las acciones y obras de los administradores de los recursos públicos.” (Ley Orgánica CGE, 2002, pág. 3).

“Art. 19.- Examen Especial.- Como parte de la auditoría gubernamental el examen especial verificará, estudiará y evaluará aspectos limitados o de una parte de las actividades relativas a la gestión financiera, administrativa, operativa y medio ambiental, con posterioridad a su ejecución, aplicará las técnicas y procedimientos de auditoría, de la ingeniería o afines, o de las disciplinas específicas, de acuerdo con la materia de examen y formulará el correspondiente informe que deberá contener comentarios, conclusiones y recomendaciones.” (Ley Orgánica CGE, 2002, pág. 3).

“Art. 26.- Informes de auditoría y su aprobación.- Los informes de auditoría gubernamental, en sus diferentes clases y modalidades, tendrán el contenido que establezcan las normas de auditoría y más regulaciones de esta Ley, incluyendo la opinión de los auditores, cuando corresponda, y la referencia al período examinado. Estos informes serán tramitados desde la emisión de la orden de trabajo de la auditoría, hasta la aprobación del informe en el término máximo de ciento ochenta días improrrogables. Los informes, luego de suscritos por el director de la unidad administrativa pertinente, serán aprobados por el Contralor General o su delegado en el término máximo de treinta días improrrogables y

serán enviados a las máximas autoridades de las instituciones del Estado examinadas de manera inmediata.” (Ley Orgánica CGE, 2002, pág. 4).

Según las leyes y normas procesales del Ecuador, los “**Días término**” son interpretados como el periodo de tiempo en el que se contabilizan solo los días hábiles, y los “**Días Plazo**” corresponden al periodo de tiempo que está compuesto por días hábiles y no hábiles.

Administrador Funcional, es el funcionario responsable del manejo integral del funcionamiento de un sistema y de los procesos automatizados en el mismo, y de toda la información ingresada, procesada u obtenida de dicho sistema.

Metodología de la Investigación

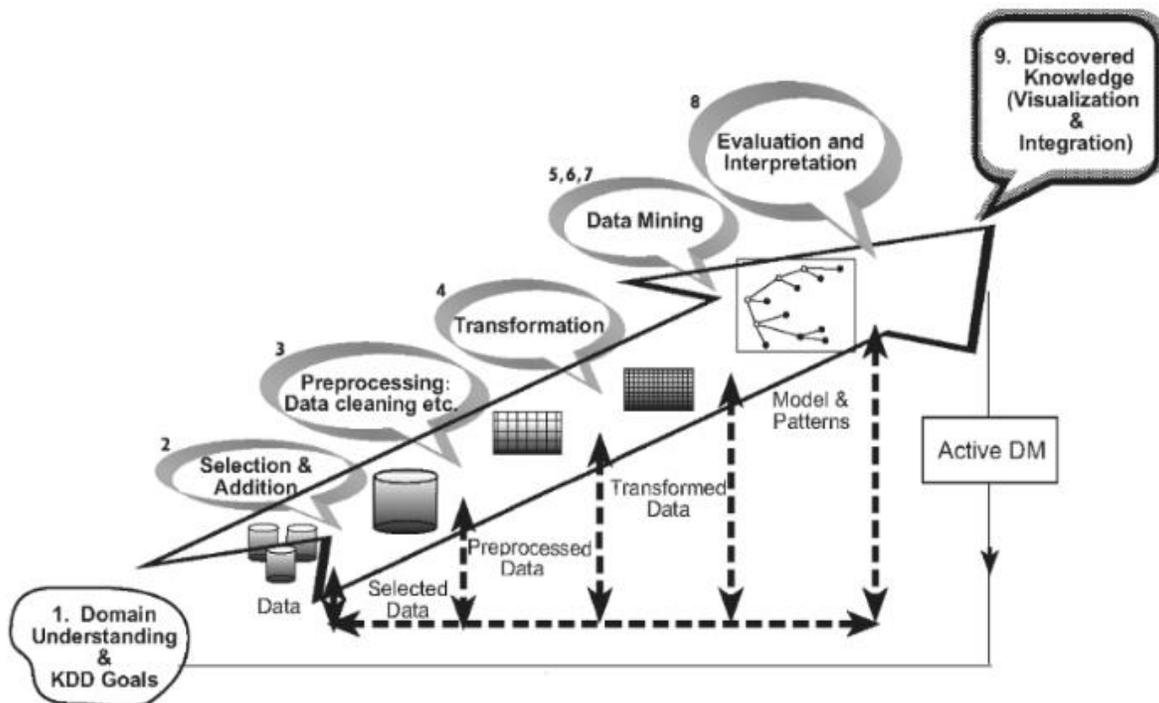
El presente estudio estará orientado por la metodología de Knowledge Discovery in Databases (KDD). Esta metodología fue propuesta por (Fayyad, Gregory Piatetsky-Shapiro, & Padhraic Smyth, 1996), consiste en ejecutar técnicas que mediante la combinación de datos podemos extraer patrones en forma de reglas o métodos matemáticos, para que el usuario los analice y se facilite la toma de decisiones.

Estas fases son: Selección, Preprocesamiento, Transformación, Minería de datos y Evaluación e interpretación,

El esquema de las cinco fases de la metodología KDD se muestra en la Figura 2.

Figura 2

Representación del Descubrimiento de Conocimiento en Base de Datos. (Oded & Lior, 2010)



Es un proceso iterativo e interactivo, involucra numerosos pasos con muchas decisiones tomadas por el usuario. Javier Landa da una visión práctica del proceso KDD (Landa, 2016), enfatizando la naturaleza interactiva del proceso. Aquí, describimos algunos de sus pasos básicos:

Realizar una revisión y un entendimiento de los datos y de la aplicación de minería de datos para tener un dominio de los conocimientos previos con el fin de identificar la meta del proceso de KDD, teniendo en cuenta el punto de vista del cliente o la necesidad de la organización.

Seleccionar un conjunto de datos, que agrupen un subconjunto de variables o datos de muestras, ya que con estos datos se llevará a cabo el descubrimiento.

Realizar la limpieza y preprocesamiento en el conjunto de datos seleccionado, aquí se ejecutan operaciones básicas que incluyen la eliminación de ruido de campos de datos vacíos, entre otros.

Buscar una herramienta que permita ejecutar una reducción de datos y hacer una proyección, para lo cual se realiza una búsqueda de características útiles en función del objetivo de la tarea. (Reducción de la magnitud)

Determinar el objetivo del KDD, respecto al uso de un método de minería de datos.

Realizar el análisis de prueba respecto a la hipótesis y al modelo de minería de datos seleccionado, éste algoritmo de minería de datos es utilizado para encontrar los patrones de datos.

Se procede a ejecutar el modelo de minería de datos, se identifican patrones de interés en el conjunto de datos seleccionado o en un subconjunto de ellos.

En este punto se procede a la interpretación de los patrones minados, posiblemente podríamos regresar a cualquiera de los pasos anteriores para más iteraciones, en este paso, se puede visualizar los patrones y modelos extraídos o tener la visualización de los datos que representan extrayendo modelos.

Finalmente, se actúa sobre el conocimiento descubierto, incorporando el conocimiento en otro sistema para la adopción de nuevas medidas o, se realiza directamente la documentación y presentación de informes a las partes interesadas, este proceso también realiza la comprobación y la solución de posibles conflictos que crean al extraer los conocimientos.

Herramientas de Minería de Datos más usadas

Dentro del análisis a las herramientas de minería de datos más usadas, se realizó una comparación de ellas y se seleccionaron las siguientes: KNIME, WEKA, Orange, RapidMiner y SAS, también se ha podido observar que los usuarios usan más de una, y que además combinan las herramientas y así encuentran puntos fuertes y diferentes (IONOS, 2022).

RapidMiner

Es una herramienta de minería de datos, que posee tres módulos, Studio, Server y Radoop, los cuales son especializados en ejecutar distintos algoritmos de la minería de datos,

cuenta con una versión gratuita y de pago para todos sus tres módulos, es una herramienta de data mining que no requiere de mucho conocimiento en programación para operar, utiliza el lenguaje de programación Java, predice comportamientos futuros usando la función predictiva de minería de datos y es la herramienta mas utilizada por las startups actualmente, a través de ella se pueden lograr algunos de los objetivos de la minería de datos, entre estos tenemos:

- Análisis de sentimientos.
- Integración de datos, tomadas de diversas fuentes de datos.
- Una visualización más clara de los resultados.
- Limpieza y preparación de los datos, mediante la transformación y detección de valores anómalos.
- Se logra un procesamiento efectivo y rápido, mediante la optimización de los datos.

(Flores, 2021)

Weka

Este software de nuestra guía de herramientas de minería de datos, es una herramienta de código abierto que surge a mediados de los años noventa, que usa el lenguaje de programación Java y ofrece conexión a bases de datos con SQL, es gratuito, el análisis de clústeres en minería de datos no es su punto fuerte, es eficiente cuando analiza cantidades pequeñas de datos.

Esta herramienta de minería de datos, ejecuta algoritmos como:

- Análisis de correlación, de clústeres o de regresión.
- Clasificación de datos, usando redes de neuronas artificiales, árboles de decisión y los algoritmos ID3 o C4.5.

(Flores, 2021)

Orange

Es una herramienta de minería de datos, que utiliza los lenguajes de programación C++ y Python, para su manejo encontramos diversos tutoriales en línea sobre el cómo utilizar este programa de minería de datos, lo cual facilita su aprendizaje, usa gráficos amigables como diagrama de dispersión, gráficos de barra, mapas de calor, y arboles de decisión, fáciles y para comprender los resultados.

además, este software tiene la función de aprender de las preferencias de quién la usa, adaptándose para proporcionar una mejor experiencia.

Este software de minería de datos ayuda a concretar las siguientes características:

- Antes de procesar los datos, permite una visualizarlos de manera atractiva y simple.
- Permite realizar clasificación, regresión y clustering de datos.

(Flores, 2021)

Knime

Esta herramienta de minería de datos, utiliza el lenguaje de programación Java y está desarrollado sobre la plataforma Eclipse, cuenta con dos versiones, una gratuita y una de pago, es una de las herramientas más populares entre los programadores y científicos de datos, analiza datos de forma integrada, descubre estructuras ocultas de información, el software de pago tiene funciones adicionales que se pueden aprovechar mejor los beneficios de la minería de datos.

Las características principales del software minería de datos KNIME podemos citar:

- Aplicaciones para procesar datos con más de 1,000 funciones y paquetes.
- Tiene tratamiento previo de los datos, como la extracción, transformación y carga.
- Este software, tiene un enfoque en el flujo de datos.

(Flores, 2021)

Sas

Esta herramienta de minería de datos es la preferida del sector de negocios, presenta una visualización interactiva de los datos, su interfaz gráfica que presenta al usuario ayuda en la interpretación para quienes no tienen muchos conocimientos en minería de datos, este software es considerado el ideal para las grandes organizaciones.

Entre las características principales de esta herramienta de minería de datos, se puede citar que la más usada Enel sector de los negocios, permite las siguientes acciones:

- Visualización de pronósticos con tecnología avanzada.
- Escalabilidad en los recursos de software que puede aumentar su eficiencia de manera progresiva.
- Software de pago

(Flores, 2021)

Comparación de herraminetas de data mining

Luego de la anterior descripción detallada de los diferentes tipos de herramientas de minería de datos, a continuación, se presenta la información de las características más relevantes de comparación en la tabla 4:

Tabla 4

Comparación de Herramientas de Software para Minería de Datos

Herramienta	Principales características	Tipo de lenguaje de programación	Tipo de Sistema operativo	Tipo de Licencia
RapidMiner	Ideal para todos los procesos, se destaca en el análisis predictivo	Java	Windows, macOS, Linux	Freeware, y diferentes versiones de pago
WEKA	Posee muchos métodos de clasificación	Java	Windows, macOS, Linux	Software libre de tipo (GPL)
Orange	Crea y presenta una visualización de datos atractiva sin que se requieran muchos conocimientos previos de programación	El núcleo del software: C++, la ampliación y lenguaje de entrada: Python	Windows, macOS, Linux	Software libre, de tipo (GPL)
KNIME	Software de minería de datos de código abierto que ha democratizado el acceso a los análisis predictivos	Java	Windows, macOS, Linux	Software libre, de tipo (GPL) a partir de la versión 2.1
SAS	Software potente, usado por grandes empresas.	Lenguaje SAS	Windows, macOS, Linux	Es freeware limitado a instituciones públicas, hay que

Software utilizado

Las herramientas de software de minería de datos utilizada para este proyecto de investigación son KNIME y WEKA, se seleccionaron a través de un benchmarking, se consideraron parámetros como el tipo de software (libre o pagado), las plataformas que soportan, los algoritmos en los que se especializan, la disponibilidad de foros de ayuda oficiales que se encuentren activos, si cuentan con una interfaz gráfica que facilite el desarrollo y por último si cuentan con una herramienta de pre visualización de resultados. Las dos herramientas están dentro de las cinco más usadas en el 2022.

El software KNIME son las siglas de Konstanz Information Miner, está a disposición de los usuarios como una herramienta de código abierto, en un inicio fue creado con fines comerciales, está escrito en Java y desarrollado sobre Eclipse, Knime está a consideración de la comunidad internacional de programadores, comparada con otras herramientas de minería de datos posee una amplia gama de funciones y paquetes y se pueden ampliar sus módulos con otras soluciones adicionales de pago (IONOS, 2022).

Las funciones predominantes del software Knime es el análisis integrativo entre los procedimientos de aprendizaje automático y de minería de datos. Además, esta herramienta tiene un excelente análisis y tratamiento previo de los datos, haciendo extracción, transformación y carga, como sus módulos son segmentados, sirve como software de minería de datos orientado al flujo de datos (IONOS, 2022).

Se ha observado que la herramienta Knime es usada en el sector de la investigación farmacéutica, financiero, se usa frecuentemente en el campo de la inteligencia empresarial (BI). Así mismo, la herramienta Knime por ser amigable y por su forma gráfica de presentar los resultados es muy usada por usuarios que se inician en el campo de la minería de datos, al poseer una amplia gama de funciones.

Finalmente es necesario citar que la herramienta Knime existe en versión gratuita y de pago (IONOS, 2022).

WEKA (Waikato Environment for Knowledge Analysis), es una herramienta desarrollada en la primera mitad de los años noventa, es un software de código abierto, desarrollado en Java, es compatible con los sistemas operativos Windows, macOS y Linux, ofrece conexión con bases de datos SQL, su interfaz gráfica es amigable y de fácil acceso al usuario, su punto fuerte dentro de la minería de datos es la clasificación de datos, tiene funciones de aprendizaje automático, también complementa tareas tan relevantes de la minería de datos como el análisis de clústeres, de correlación o de regresión, usa redes de neuronas artificiales, árboles de decisión y algoritmos ID3 o C4.5. (IONOS, 2022).

siendo capaz de procesar en ellas los datos solicitados. Así mismo, presenta un sinnúmero de funciones de aprendizaje automático y complementa tareas tan relevantes de la minería de datos como el análisis de clústeres, de correlación o de regresión, así como la clasificación de datos, este último es un punto fuerte del software minería, al usar redes de neuronas artificiales, árboles de decisión y algoritmos C4.5 o ID3. (IONOS, 2022).

Weka recibió el premio en reconocimiento “SIGKDD Service Award” en el año 2005 de la Association for Computing Machinery, por su contribución a la investigación, comparada con otras herramientas de minería de datos, se destaca en su contribución en la investigación y la enseñanza. (IONOS, 2022).

Hardware utilizado

Para ejecutar KNIME requiere una PC que permita instalar/ejecutar Linux, Windows (XP y posterior) o Mac OSX. No se necesita hardware de gráficos sofisticado, los sistemas de múltiples núcleos son una ventaja, ya que KNIME utiliza múltiples núcleos. El espacio disponible en el disco duro (NO en la memoria principal) limita la cantidad de datos procesables; se recomiendan varias decenas de GB de espacio libre. La memoria principal debe ser al menos de 1 GB o más, en sistemas de 32 bits se pueden usar hasta 1,5 GB, más en sistemas de 64 bits. Para el correcto funcionamiento de la herramienta Weka, se requiere

una PC con Linux, Windows 98SE, Me, NT, 2000, XP, Vista, Windows 7 o Windows 8.(Arquitectura Windows x86 , x64) o Mac OS X. Debe contener al menos un procesador Pentium IV 1 GHz o superior, el disco duro contendrá sobre los 132,5 Mb libres, la memoria RAM de 512 Mb o superior, la tarjeta gráfica debe ser compatible con Open GL con resolución 800x600 o superior.

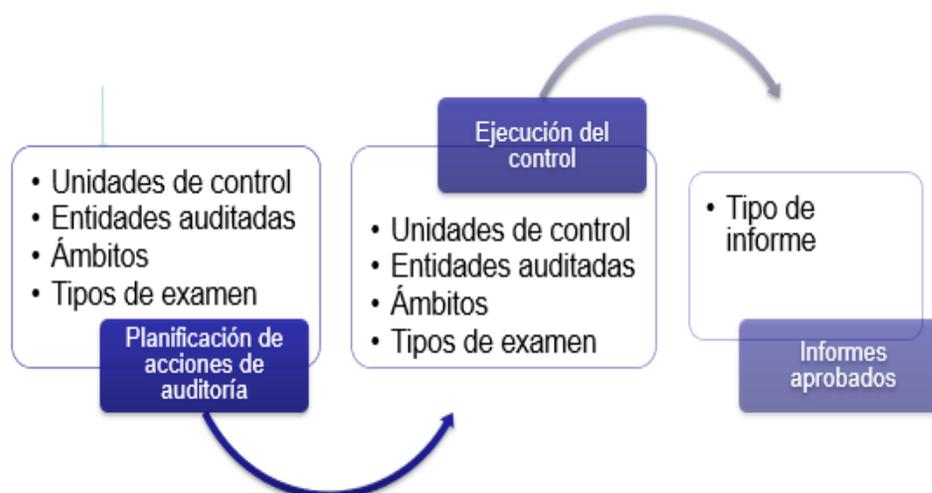
Capítulo III

Desarrollo

En este capítulo se describe el estudio realizado de acuerdo a las fases de la metodología *Knowledge Discovery in Databases* (KDD). Se realizó el levantamiento de información para conocer los procesos de control que se ejecutan para precautelar la utilización de los recursos del Estado a las diferentes entidades públicas del Estado. De los procesos identificados se han seleccionado tres: planificación de acciones de auditoría, ejecución del control (órdenes de trabajo), e informes aprobados que se visualiza en la Figura 3.

Figura 3

Mapa de Procesos seleccionados para el estudio



El primer proceso del estudio, es el de la planificación de las acciones de control de auditoría, considera la fase de revisión, generación y la aprobación del Plan Anual de Acciones de Control tanto de la auditoría externa como de la auditoría interna.

Respecto al segundo proceso seleccionado, referente a la ejecución de las ordenes de trabajo tanto para las unidades de Auditoría Interna como para las unidades de Auditoría Externa, es la fase de la ejecución y cumplimiento de las acciones de control que se cumplen

en las entidades públicas, inicia desde el instante de su ejecución hasta cuando se concluye con el trámite de la generación del informe de auditoría.

Referente al tercer proceso, está la aprobación de los informes de auditoría, que son elaborados una vez que concluye la ejecución de la orden de trabajo, es decir cuando ha sido revisada y supervisada la orden de trabajo, se genera el borrador del informe de auditoría, posterior a éste es generado el informe de auditoría, y aprobado por el director de la Unidad Administrativa o por el Subcontralor de Auditoría o el Contralor General del Estado.

Recopilación de la Información

Se aplicó el método de la observación directa, en el presente estudio no se requiere la aplicación de cuestionarios y/o encuestas. Para la recopilación de información se utilizaron dos fuentes de información física y digital.

La información en físico reposa en el archivo de la entidad de control. Los datos en formato digital se encuentran en el sistema informático de propiedad de la Entidad de Control. La documentación que se maneja en la institución es de carácter legal y reservada, y se rige en la normativa vigente del país. (Ley Orgánica CGE, 2002).

Una vez realizada la revisión y análisis de la información física se identifican tipos de exámenes que se aplican a las entidades del sector público, dependiendo los ámbitos de control, indicadores por medio de los cuales se evalúa el proceso y el cumplimiento de las acciones de control, informes que pueden estar aprobados o no, el alcance del estudio, expedientes.

En el repositorio digital se ha encontrado información de diferentes procesos que ejecuta la Entidad de Control y se busca clasificarla por años, por instituciones, unidades, tipos de exámenes, ordenes de trabajo, informes de auditoría que han sido escaneados que reflejan lo realizado respecto a la evaluación de las gestiones, administrativa, financiera, operativa, etc.

Clasificación de la información en archivo físico

En el archivo físico, se encontró la información catalogada por expediente, dentro de este constan el informe de auditoría, cierta información referente a la orden de trabajo, en otro apartado la planificación aprobada de las acciones de control, por lo que se revisa y se realiza una clasificación en base a la unidad de control ejecutora, tipo de examen, ámbitos y tipo de informe.

Los tipos de exámenes que se realizan en la entidad de control se muestran en la Tabla 5.

Tabla 5

Tipos de Examen de Auditoría

Tipo de examen	Descripción
Auditoría de aspectos ambientales	Tipo de examen que corresponde a los exámenes de evaluación de impacto ambiental en los términos establecidos en la Ley de Gestión Ambiental.
Auditoría de obras públicas	Tipo de examen que evalúa la administración de las obras en construcción, la gestión de los contratistas, el manejo de la contratación pública, la eficacia de los sistemas de mantenimiento, el cumplimiento de las estipulaciones contractuales y los resultados físicos que se obtengan en el programa o proyecto específico que está siendo sometido a examen.
Auditoría financiera	El tipo de examen de auditoría financiera, informará respecto a un período determinado, sobre la evaluación de las cifras

	<p>presentadas en los estados financieros de una institución pública, ente contable, programa o proyecto, concluye con la elaboración del informe de auditoría, en el que se incluirán las opiniones correspondientes.</p>
Evaluación control interno	<p>Tipo de examen, considerado en la Auditoría Financiera.</p> <p>Es el tipo de examen más usado ya que dentro de su ejecución, considera a los demás tipos, este examen es el que verifica, estudia y evalúa aspectos limitados o de una parte de las actividades relativas a la gestión financiera, administrativa, operativa y medio ambiental, con posterioridad a su ejecución, aplica las técnicas y procedimientos de control de auditoría, de la ingeniería o afines, o de las disciplinas específicas, de acuerdo con la materia de examen y formula el correspondiente informe que contiene comentarios, conclusiones y recomendaciones.</p>
Examen especial	
Supervisión firmas privadas	<p>Tipo de examen considerado en la Auditoría de Gestión.</p>
Declaraciones patrimoniales juradas	<p>Tipo de examen considerado en la Auditoría de Gestión.</p> <p>Examen considerado en la Auditoría de Gestión.</p>

Operativos de control

vehicular

Auditoría de gestión

Tipo de examen que está dirigido a examinar y evaluar la gestión del control interno, utiliza recursos humanos de carácter multidisciplinario, mide el desempeño de una organización, ente contable, o la ejecución de programas y proyectos, de acuerdo a principios y criterios de la economía, efectividad y eficiencia.

A continuación, se describen los tipos de unidades de control y sus respectivos ámbitos, tipos de informes, algunos de estos parámetros y campos se consideran para el estudio.

Tipos de Unidades y sus Ámbitos

Dna1 Dirección Nacional de Auditoría de Administración Central

Con sus ámbitos:

- ✓ Administrativo,
- ✓ Asuntos del Exterior,
- ✓ Asuntos Internos,
- ✓ Defensa Nacional,
- ✓ Electoral,
- ✓ Jurisdiccional,
- ✓ Legislativo, y
- ✓ Transparencia y Control Social

Dna2 Dirección Nacional de Auditoría de Sectores Sociales

Con sus ámbitos:

- ✓ Bienestar Social,
- ✓ Cultura,

- ✓ e) Deporte,
- ✓ Desarrollo Urbano y Vivienda,
- ✓ Educación,
- ✓ Laboral, y
- ✓ Turismo

Dna3 Dirección Nacional de Auditoría de Deuda Pública y Finanzas

Con sus ámbitos:

- ✓ Entidades Financieras Públicas, y
- ✓ Finanzas

Dna4 Dirección Nacional de Auditoría de Telecomunicaciones, conectividad y sectores productivos

Con sus ámbitos:

- ✓ Comunicaciones,
- ✓ Electricidad,
- ✓ Industrialización, y
- ✓ Telecomunicaciones

Dna5 Dirección Nacional de Auditoría de Gobiernos Autónomos Descentralizados

Con sus ámbitos:

- ✓ Gobiernos Autónomos Descentralizados (Regionales, Provinciales, Municipales, Distritos Metropolitanos Autónomos y Parroquias Rurales),
- ✓ Mancomunidades de Gobiernos Autónomos Descentralizados en la misma provincia,
- ✓ Entidades Asociativas de Gobiernos Autónomos Descentralizados en la misma provincia, y
- ✓ Empresas o instituciones públicas creadas o que se crearen por los Gobiernos Autónomos Descentralizados, y/o aquellas creadas por varios Gobiernos

Autónomos Descentralizados o por las Mancomunidades de Gobiernos

Autónomos Descentralizados en la misma provincia.

Dna6 Dirección Nacional de Auditoría de Recursos Naturales

Con sus ámbitos:

- ✓ Agropecuario,
- ✓ Ambiente, y
- ✓ Recursos Naturales

Dna7 Dirección Nacional de Auditoría de Salud y Seguridad Social

Con sus ámbitos:

- ✓ Salud, y
- ✓ Seguridad Social (FFAA, Policía Nacional, Instituto Ecuatoriano de Seguridad Social)

Dna8 Dirección Nacional de Auditoría de Transporte, Vialidad, Infraestructura

Portuaria y Aeroportuaria

Con sus ámbitos:

- ✓ Transporte, y
- ✓ Vialidad

Tipos de Informes

Los tipos de informes, según menciona la Constitución de la República del Ecuador, en una de las funciones de la Entidad de Control es la de “Determinar responsabilidades administrativas y civiles culposas e indicios de responsabilidad penal, relacionadas con los aspectos y gestiones sujetas a su control, sin perjuicio de las funciones que en esta materia sean propias de la Fiscalía General del Estado ecuatoriano” (Asamblea Nacional del Ecuador, 2008, pág. 44).

El resultado del análisis y estudio de la utilización de los recursos estatales de las instituciones públicas y de las personas jurídicas de derecho privado que utilizan recursos públicos son los informes de auditoría que pueden ser de tipo General o Penal.

Las responsabilidades administrativas y civiles culposas son identificadas en el **Informe de tipo General**.

Las responsabilidades con indicios de responsabilidad penal se identifican en el **Informe de tipo Penal**.

Indicadores de Control

Luego del análisis y estudio en la materia de control, se identifican algunos de los indicadores que permiten cuantificar los procesos de auditoría y que serán considerados para el presente estudio.

- Número de exámenes planificados y cumplidos
- Número de informes entregados y aprobados
- Tiempo que conlleva la acción de control desde que se aprueba una orden de trabajo hasta la aprobación del informe, son 180 días.

Clasificación de la información en archivo digital

Se realizó coordinaciones con la Unidad de Tecnologías a fin de conocer la estructura de los datos de los procesos seleccionados, accediendo al repositorio de datos de desarrollo. Se construyeron consultas para extraer la data desde el repositorio de desarrollo institucional, usando SQL Server y los datos obtenidos fueron exportados a hojas de cálculo.

Para la recopilación de la data, se digitó la información de los documentos físicos y en otros casos se revisa la información, así mismo se buscan datos registrados en los Sistemas de Planificación y de Control, se recopilan y se unifican en formato Excel.

A continuación, se muestra el proceso realizado para clasificar la información y seleccionar la muestra.

En la Figura 4 y 5 se realiza una exploración de los datos del Plan aprobado para el año

2022

Figura 4

Consulta de datos del plan aprobado del año 2022

```
--Plan Aprobado 2022
select a.accId, a.accDescripcion, a.accNombreEntidad, u.uniCodigo,e.entAmbito, u.uniNombre,c.catCodigo, e.entSector,
       c.catDescripcion,*
from Control.Accion a inner join [redacted] pu on pu.plauniId=a.plauniId
inner join PlanControl.[redacted] n on a.accIdRelacionado = a1.accId
inner join PlanControl.PlanControl pc on pc.placonId=pu.placonId and p.[redacted]
inner join PlanControl.[redacted] te on te.tipexaId = a.tipexaId
inner join Catastro.Entidad e on e.entCodigo = pu.[redacted]
inner join General.catalogo c on [redacted] Sector
inner join [redacted] u on u.uniCodigo = e.entAmbito
where tipexaEstado='A'
and a.tipexaId not in ([redacted])
```

accId	accDescripcion	accNombreEntidad	uniCodigo	entAmbito	uniNombre	catCodigo	entSector	catDescripcion	accId	accIdRelacion
1	las etapas de determinación, recaudación, depósito y registro d...	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
2	al cumplimiento de las recomendaciones contenidas en los infor...	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
3	los procesos de adquisición de bienes y contratación de servi...	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
4	OPERATIVO DE CONTROL VEHICULAR - CARNAVAL	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
5	Operativo de control vehicular	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
6	OPERATIVO DE CONTROL VEHICULAR - DÍA DE DIFUNTOS ...	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
7	OPERATIVO DE CONTROL VEHICULAR - NAVIDAD	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
8	OPERATIVO DE CONTROL VEHICULAR - VIERNES SANTO	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
9	OPERATIVO DE CONTROL VEHICULAR - DÍA DEL TRABAJO	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]
10	OPERATIVO DE CONTROL VEHICULAR - BATALLA DE PICH...	CONSEJO NACIONAL ELECTORAL	2200	2200	DIRECCIÓN DE AUDITORIA 1	[redacted]	[redacted]	Elector	[redacted]	[redacted]

Figura 5

Consulta de datos del plan aprobado del año 2022

```
--Plan Aprobado 2022
select a.accId, a.accDescripcion, a.accNombreEntidad, u.uniCodigo,e.entAmbito, u.uniNombre,c.catCodigo, e.entSector,
       c.catDescripcion,*
from Control.Accion a inner join PlanControl.PlanUnidad pu on pu.plauniId=a.plauniId
inner join PlanControl.Accion a1 on a.accIdRelacionado = a1.accId
inner join PlanControl.PlanControl pc on pc.placonId=pu.placonId and pc.placonId=8
inner join PlanControl.TipoExamen te on te.tipexaId = a.tipexaId
inner join Catastro.Entidad e on e.entCodigo = pu.plauniCodigoEntidad
inner join General.catalogo c on c.catCodigo = e.entSector
inner join Talento.Unidad u on u.uniCodigo = e.entAmbito
where tipexaEstado='A'
and a.tipexaId not in (1,17,11,20)
```

catCodigo	TipoOCV	accEjercicioEconomico	accTipoDisposicion	accDisposicionDetalle	accplauniIdOrigen	accplanuniEstado	pedauid	accDiasPrevioInforme	plauniId	placonId	plauniCo...	plauniDescripcionEntidad
1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	15	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
2	NULL	NULL	NULL	NULL	NULL	NULL	NULL	15	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
3	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
4	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
5	4039	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
6	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
7	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
8	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
9	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,
10	4038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2308	8	11959	TRIBUNAL SUPREMO ELECTOR,

En la Figura 6 tenemos la preparación de la consulta para encontrar los datos de la ejecución de las acciones de control.

Figura 6

Consulta de la Ejecución de las acciones de control según el año 2022

```
--Ejecución del Control del año 2022
select o.ordtraId, a1.accId,pu.plauniId,a.accId, a.accDescripcion, a.accNombreEntidad,
a.accFechaExaminarDesde, a.accFechaExaminarHasta, ordtraNumeroOrdenTrabajo, ordtraFechaOrdenTrabajo, *
from Control.Accion a inner join PlanControl.PlanUnidad pu on pu.plauniId=a.plauniId
inner join PlanControl.Accion a1 on a.accIdRelacionado = a1.accId
inner join PlanControl.PlanControl pc on pc.placonId=pu.placonId and pc.placonId=8
inner join PlanControl.TipoExamen te on te.tipexaId = a.tipexaId
inner join Control.OrdenTrabajo o on o.accId=a.accId |
where tipexaEstado='A'
and a.tipexaId not in (1,11,17,20)
```

ordtraId	accId	plauniId	accId	accDescripcion	accNombreEntidad	accFechaExaminarDesde	accFechaExaminarHasta	ordtraNumeroOrdenTrabajo	
252	27779	431...	2238	27251	A los procesos preparatorio, precontractual, contractual, ejec...	CORPORACIÓN NACIONAL DE FINA...	2021-01-01 00:00:00.000	2022-08-31 00:00:00.000	0003-DNA3-CONAFIPS-...
253	26743	431...	2238	27252	Al cumplimiento de las recomendaciones contenidas en los inf...	CORPORACIÓN NACIONAL DE FINA...	2020-07-01 00:00:00.000	2022-04-30 00:00:00.000	0002-DNA3-CONAFIPS-...
254	25227	431...	2240	27253	a la adquisición, existencia, control, mantenimiento y custodia...	GOBIERNO AUTÓNOMO DESCENTR...	2017-01-01 00:00:00.000	2021-12-31 00:00:00.000	0001-DPE-GADPE-AI-202...
255	26858	431...	2240	27254	a las etapas preparatorias, precontractuales, contractuales, d...	GOBIERNO AUTÓNOMO DESCENTR...	2017-01-01 00:00:00.000	2021-12-31 00:00:00.000	0002-DPE-GADPE-AI-202...
256	25146	431...	2241	27255	a las fases precontractual, contractual, ejecución, liquidación ...	MINISTERIO DE ECONOMÍA Y FINAN...	2018-01-01 00:00:00.000	2020-12-31 00:00:00.000	0001-DNA3-MEF-AI-2022
257	25311	431...	2242	27256	A la administración de la infraestructura tecnológica y segurid...	BANCO CENTRAL DEL ECUADOR (Q...	2018-01-01 00:00:00.000	2021-12-31 00:00:00.000	0001-DNA3-BCE-AI-2022
258	25312	431...	2242	27257	A los procesos de operaciones propias del Banco Central del ...	BANCO CENTRAL DEL ECUADOR (Q...	2017-01-01 00:00:00.000	2021-12-31 00:00:00.000	0002-DNA3-BCE-AI-2022
259	26827	431...	2242	27258	A la administración del parque automotor del Banco Central d...	BANCO CENTRAL DEL ECUADOR (Q...	2019-05-01 00:00:00.000	2022-04-30 00:00:00.000	0004-DNA3-BCE-AI-2022
260	27865	431...	2242	27259	Al cumplimiento de las recomendaciones constantes en los inf...	BANCO CENTRAL DEL ECUADOR (Q...	2021-07-01 00:00:00.000	2022-08-31 00:00:00.000	0005-DNA3-BCE-AI-2022
261	26741	431...	2241	27260	a la administración de los Inventarios de Bienes de Uso y Con...	MINISTERIO DE ECONOMÍA Y FINAN...	2017-01-01 00:00:00.000	2021-12-31 00:00:00.000	0002-DNA3-MEF-AI-2022
262	27854	431...	2241	27261	al cumplimiento de las recomendaciones constantes en los inf...	MINISTERIO DE ECONOMÍA Y FINAN...	2021-07-01 00:00:00.000	2022-08-31 00:00:00.000	0003-DNA3-MEF-AI-2022
263	25096	431...	2244	27262	a las fases preparatoria y precontractual del proceso de Dele...	DIRECCIÓN GENERAL DE AVIACIÓN ...	2019-05-01 00:00:00.000	2021-07-31 00:00:00.000	0001-DNA8-TVIFYA-DGA...

En la Figura 7 tenemos la preparación de la consulta para encontrar los datos de los informes aprobados de las acciones planificadas en el año 2022.

Figura 7

Consulta de datos de Informes Aprobados del Año 2022

```
-- Informes aprobados en el año 2022 según la planificación del año 2022
select o.[ordtraId] ,o.[accId],o.[detperIdRegistra],o.[ordtraFechaRegistro],o.[catCodigoEstado]
,o.[ordtraNumeroOrdenTrabajo],o.[ordtraObjetivos],o.[ordtraAlcance],o.[ordtraEntidadAuditora]
,o.[ordtraUnidadAuditora],o.[ordtraEntidadExaminada],o.[proId],o.[canId],o.[ciuId],o.[ordtraLugarAccionControl]
,o.[ordtraFechaOrdenTrabajo],o.[detperIdFirmante],o.[tipexaId],o.[ordtraPeriodoExaminarDesde],
o.[ordtraPeriodoExaminarHasta],o.[ordtraTotalDiasExamen],o.[ordtraDiasHombreExamen],o.[ordtraDiasHombreRevision],
o.[denautId],o.[catCodigoTipoAuditoria] ,o.[catCodigoTipoOCV],o.[autconId],o.[ordtraAsignacionOCV],
o.[ordtraFechaOCVDesde],o.[ordtraFechaOCVHasta],o.[ordtraHorarioOCVDesde],o.[ordtraHorarioOCVHasta],
o.[ordtraFechaControlPDesde],o.[ordtraFechaControlPHasta],o.[catCodigoArticuloEntidad],o.[ordtraEntidadesAdscritas],
o.[ordtraEntidadesPublicas],o.[ordtraEjercicioEconomico],o.[ordtraEntidadesRelacionadas],o.[ordtraEntiRelaNovedad],
o.[ordtraNumeroOrdenTrabajoDocumentos],o.[ordtraIdentificadorSecuencial],o.[ordtraUnidadAsignada],i.[infId],
i.[ordtraId],i.[detperIdRegistra],i.[apronId],i.[infNumeroInforme],i.[infFechaAprobacion] ,i.[infAprobador],
i.[infCargoAprobador],i.[infdetperIdAprobador] ,i.[infTipoInforme],i.[infNumeroPaginas],i.[infIntroduccion],
i.[infLimitacionAlcance],i.[infNIS],i.[infAnioNIS] ,i.[infEstado],i.[infFechaRegistro],i.[infMontoIndeterminado],
i.[infUnidadControl],i.[infAutorizacion],i.[infAlcance],i.[infUnidadControlHistorico],i.[infSeguiUnidad],
i.[infEntidadExaminada],i.[infNumeroOrdenTrabajo],i.[infFechaOrdenTrabajo],i.[infUnidadSeguimiento],
i.[infEstadoSegui],i.[infSinResponsabilidad],i.[infPeriodoExaminarDesde],i.[infPeriodoExaminarHasta]]
from Control.OrdenTrabajo o
inner join Control.Informe i on o.ordtraId = i.ordtraId
where year(o.ordtraFechaRegistro)=2022
and year(i.infFechaAprobacion)=2022
and i.infTipoInforme='G'
```

ordtraNumeroOrdenTrabajo	infId	ordtraId	detperIdRegistra	apronId	infNumeroInforme	infFechaAprobacion	infAprobador	infCargoAprobador
0001-OP1-GADMI-AI-2022	22028	25100	49586	NULL	DP1-0004-2022	2022-04-08 00:00:00.000	TORO ECHEVERRIA PAUL ALEXANDER	DIRECTOR/A PROVINCIAL 2 (Titular)
0001-DNA5-GAD-2022-EEP	22035	25476	52123	NULL	DNA5-GAD-0025-2022	2022-04-08 00:00:00.000	RIOFRIO GONZÁLEZ CARLOS ALBERTO	CONTRALOR GENERAL DEL ESTADO
0001-OP1-EP-EMAPA-A-AI-2022	22065	25237	51812	NULL	DPT-EP-EMAPA-A-AI-0001-2022	2022-04-20 00:00:00.000	RIVERA PIÑALOZA DARWIN VLADIMIR	DIRECTOR/A PROVINCIAL 2 (Titular)
0001-OP1-UTN-AI-2022	22073	25124	49586	NULL	DP1-0006-2022	2022-04-22 00:00:00.000	TORO ECHEVERRIA PAUL ALEXANDER	DIRECTOR/A PROVINCIAL 2 (Titular)
0001-OP1-GADMACO-AI-2022	22099	25412	49586	NULL	DP1-0009-2022	2022-04-28 00:00:00.000	TORO ECHEVERRIA PAUL ALEXANDER	DIRECTOR/A PROVINCIAL 2 (Titular)
0001-OPA-GADMCC-AI-2022	22100	25186	52571	NULL	DPA-0021-2022	2022-04-28 00:00:00.000	GONZALEZ BARROS JUAN FERNANDO	DIRECTOR/A PROVINCIAL 2 (Encargado)
0001-DNA2-MD-AI-2022	22104	25157	51076	NULL	DNA2-0019-2022	2022-04-29 00:00:00.000	ÁLVAREZ GARCÉS ZONNIA YOLANDA	DIRECTOR/A NACIONAL DE AUDITOF
0001-DNA2-EPN-AI-2022	22109	25137	51076	NULL	DNA2-0020-2022	2022-05-03 00:00:00.000	ÁLVAREZ GARCÉS ZONNIA YOLANDA	DIRECTOR/A NACIONAL DE AUDITOF
0001-DPGY-BCBG-AI-2022	22110	25251	52056	NULL	DPGY-0034-2022	2022-04-27 00:00:00.000	FLORES ARROYO XAVIER GASTÓN	DIRECTOR PROVINCIAL 2 (Titular)
0001-DNA6-RN-RNIAP-AI-2022	22122	25250	52778	NULL	DNA6-RN-0027-2022	2022-05-04 00:00:00.000	LOAIZA GRANIZO EDITH JOVITA DEL CISNE	DIRECTOR/A NACIONAL DE AUDITOF
0001-DNA2-MIES-AI-2022	22123	25284	51076	NULL	DNA2-0022-2022	2022-05-06 00:00:00.000	ÁLVAREZ GARCÉS ZONNIA YOLANDA	DIRECTOR/A NACIONAL DE AUDITOF

En la Figura 8 tenemos la preparación de la consulta para encontrar los datos de los informes aprobados, analizando los tiempos de ejecución, considerando la fecha de inicio de la orden de trabajo hasta la aprobación del informe de tipo General, según los diferentes ámbitos de control.

Figura 8

Consulta de los Informes Aprobados del Año 2022 con el Tiempo en días.

```

--Tiempo de ejecución de la orden de trabajo y la aprobación del Informe
-- encontrar los mejores tiempos de ejecución de la orden de trabajo
--según los ámbitos
select i.infNumeroInforme, c.catDescripcion, e.entNombre, u.uniNombre,
[[Control].[fnDiasLaboralesFeriadosNacionalesLocales](o.ordtraFechaRegistro,i.infFechaAprobacion)] TotalDias,
case
  when o.tipexaId='5' then 'Auditoria de Gestión'
  when o.tipexaId='8' then 'Auditoria de Obras Públicas'
  when o.tipexaId='9' then 'Auditoria Financiera'
  when o.tipexaId='13' then 'Examen Especial'
  when o.tipexaId='15' then 'Examen Especial de Ingeniería'
  when o.tipexaId='19' then 'Supervisión Firmas Privadas'
  when o.tipexaId='21' then 'Declaraciones Patrimoniales Juradas'
  when o.tipexaId='22' then 'Operativos de Control Vehicular'
  when o.tipexaId='23' then 'Examen Especial de Paraísos Fiscales'
  when o.tipexaId='24' then 'Auditoria de Gestión'
end tipoExamen
from Control.OrdenTrabajo o
inner join Control.Informe i on o.ordtraId = i.ordtraId
inner join Catastro.Intidad e on e.entCodigo=o.ordtraIntidadExaminada
inner join General.catalogo c on c.catCodigo = e.entCodigo
inner join Talento.Unidad u on u.uniCodigo=e.entAmbito
where year(ordtraFechaRegistro)=2022
and year(InfFechaAprobacion)=2022
and i.infTipoInforme = 'G'

```

infNumeroInforme	catDescripcion	entNombre	uniNombre	TotalDias	tipoExamen
DNA1-0086-2022	Administrativo	SECRETARIA TÉCNICA DE GESTIÓN INMOBILIARIA DEL S...	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	160	Examen Especial
DNA1-0079-2022	Administrativo	SECRETARIA TÉCNICA DE GESTIÓN INMOBILIARIA DEL S...	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	150	Examen Especial
DNA1-0072-2022	Administrativo	PRESIDENCIA DE LA REPÚBLICA	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	144	Examen Especial
DPM-0025-2022	Administrativo	SECRETARIA NACIONAL DE PLANIFICACIÓN	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	167	Examen Especial
DNA1-0052-2022	Administrativo	DIRECCIÓN GENERAL DE REGISTRO CIVIL, IDENTIFICACI...	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	178	Examen Especial
DNA1-0087-2022	Administrativo	PROCURADURÍA GENERAL DEL ESTADO	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	170	Examen Especial
DNA1-0093-2022	Administrativo	SERVICIO NACIONAL DE CONTRATACIÓN PÚBLICA	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	179	Examen Especial
DNA1-0062-2022	Administrativo	SERVICIO ECUATORIANO DE NORMALIZACIÓN	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	121	Examen Especial
DNA1-0069-2022	Administrativo	INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC)	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	145	Examen Especial
DNA1-0097-2022	Administrativo	INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC)	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	93	Examen Especial
DNA1-0083-2022	Administrativo	INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC)	DNA 1 - DIRECCIÓN NACIONAL DE AUDITORIA DE ADMIN...	160	Examen Especial

En la Figura 9 tenemos la preparación de la consulta con el fin de encontrar los datos del tiempo de la ejecución de la orden de trabajo, tiempo de aprobación del informe y aprobación del informe.

Figura 9

Consulta de los Tiempos de Ejecución de la Orden de Trabajo, Tiempos de Aprobación del Informe

```

--tiempo de ejecución de la orden de trabajo hasta la aprobación del informe,
--tiempo de ejecución de la orden de trabajo
--tiempo de aprobación del informe
select infNumeroInforme,
[Control].[fnDiasLaboralesFeriadosNacionalesLocales](o.ordtraFechaRegistro,i.infFechaAprobacion) TotalDias,
[Control].[fnDiasLaboralesFeriadosNacionalesLocales](o.ordtraFechaRegistro,cc.concalFechaRegistro) DiasOT,
[Control].[fnDiasLaboralesFeriadosNacionalesLocales](cc.concalFechaRegistro,i.infFechaAprobacion) DiasAprob
from Control.OrdenTrabajo o
inner join Control.Informe i on o.ordtraId = i.ordtraId
inner join Control.ControlCalidad cc on o.ordtraId = cc.ordtraId
where year(ordtraFechaRegistro)=2022
and year(infFechaAprobacion)=2022

```

infNumeroInforme	TotalDias	DiasOT	DiasAprob	
1	DPI-0015-2022	108	52	57
2	DPI-0016-2022	108	55	54
3	DPGY-0034-2022	80	56	25
4	DPT-0009-2022	116	58	59
5	DPT-0008-2022	108	58	51
6	DPI-0004-2022	68	58	11
7	DPGY-0035-2022	90	65	26
8	DPI-0014-2022	98	65	34
9	DPI-0022-2022	128	65	64
10	DPE-0009-2022	90	65	26

Selección de conjuntos de datos

De los resultados de las consultas en sql ejecutadas se copian en hojas de Excel, se analizó los datos, se realizó la limpieza y selección del conjunto de datos para el estudio, Se seleccionó los datos por tipo examen, ámbitos de control y tipo informe.

En la Figura 10, se observa la clasificación y unificación de los datos de la planificación aprobada del año 2022, seleccionados para el estudio, el resultado tenemos en el archivo en Excel "Planificacion_Vs_01_2022.xlsx".

Figura 10

Datos seleccionados del Proceso de Planificación del Año 2022.

accId	Accion Descripcion	Nombre Entidad	FechaExaminarD	FechaExaminarH	FechaRegist	DiasHombre	DiasHombre	plauNiDescripcionEntidad	plauNiDescri	plauNiTipo	Descripcion	Nombr
29528	Commemoración del Día de la	DIRECCIÓN PROVINCIAL DE	2022-02-11 00:00	2022-02-14 00:00	2022-02-09 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Al proceso d	GOBIEF
29531	mediante la constatación físico	DIRECCIÓN PROVINCIAL DE	2022-02-10 00:00	2022-02-15 00:00	2022-02-09 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Examen espi	NULL
29533	Control vehicular por el Día de	DIRECCIÓN PROVINCIAL DE	2022-02-10 00:00	2022-02-13 00:00	2022-02-10 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		A los proces	GOBIEF
29597	operativo de control vehicular	DIRECCIÓN PROVINCIAL DE	2022-04-08 00:00	2022-04-11 00:00	2022-04-08 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Examen espi	NULL
29637	CONTROL VEHICULAR - PROVH	DIRECCIÓN PROVINCIAL DE	2022-04-22 00:00	2022-04-24 00:00	2022-04-20 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		A los ingres	GOBIEF
29643	Control vehicular por el Feri	DIRECCIÓN PROVINCIAL DE	2022-04-28 00:00	2022-04-28 00:00	2022-04-28 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		A los proces	GOBIEF
29818	Fundación de Guayaquil	DIRECCIÓN PROVINCIAL DE	2022-07-22 00:00	2022-07-25 00:00	2022-07-22 0	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Al proceso d	GOBIEF
29832	Precautelar el buen uso de los	DIRECCIÓN PROVINCIAL DE	2022-07-29 00:00	2022-07-29 00:00	2022-08-01 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Examen espi	NULL
29945	En cumplimiento a lo dispuest	DIRECCIÓN PROVINCIAL DE	2022-10-06 00:00	2022-10-07 00:00	2022-10-06 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Examen espi	NULL
29946	En cumplimiento a lo dispuest	DIRECCIÓN PROVINCIAL DE	2022-10-06 00:00	2022-10-06 00:00	2022-10-06 1	80	1	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		A la recuper	GOBIEF
29586	a la auditoria a los Estados Fin	FUNDACIÓN TERMINAL TEF	2018-01-01 00:00	2020-12-31 00:00	2022-04-04 1	16	3	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE			NULL
29778	por los ejercicios económicos	GOBIERNO AUTÓNOMO DE	2019-01-01 00:00	2020-12-31 00:00	2022-07-04 1	15	3	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		A la formula	EMPRE
29934	auditar la administración del c	MUNICIPIO DE LOJA	2021-01-01 00:00	2021-12-31 00:00	2022-10-03 0	3	3	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Al proceso y	NULL
29955	Por los ejercicios económicos	REGISTRO MUNICIPAL DE L	2017-01-01 00:00	2021-12-31 00:00	2022-10-13 1	9	3	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		AL PROCESO	EMPRE
29545	al Proyecto de Irrigación Tecn	MINISTERIO DE AGRICULTU	2017-01-01 00:00	2021-12-31 00:00	2022-02-25 0	130	13	CONTRALORIA GENERAL DEL I	DNA 6 - DIRE	AE	Al proceso d	NULL
29905	A los procesos preparatorio, p	EMPRESA PÚBLICA DEL AGL	2017-01-01 00:00	2021-12-31 00:00	2022-09-08 1	324	20	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		Al proceso d	GOBIEF
29584	A ingresos y gastos	FEDERACIÓN ECUATORIAN	2017-01-01 00:00	2021-12-31 00:00	2022-04-04 0	161	20	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		control vehi	GOBIEF
29729	a la ejecución, liquidación y pi	EMPRESA PÚBLICA MUNICI	2017-01-01 00:00	2022-04-30 00:00	2022-05-31 1	104	20	CONTRALORIA GENERAL DEL I	AUDITORÍA E AE		a la administ	BANCC

En la Figura 11, se muestra la clasificación y unificación de los datos de la ejecución de las acciones de control hasta la aprobación del informe.

Figura 11

Datos seleccionados de la Ejecución del Control del año 2022

ordtraid	ordtraFechaRegistro	NumeroOrden Trabajo	Objetivos	Alcance	EntidadAudi	UnidadAudit	EntidadExan	LugarAccion	FechaOrden	detperidFirm	tipexald	PeriodoExan	PeriodoExan
25086	2022-01-03 11:10:52.1	0001-DPI-AE-2022	Los objetivos: a los gastos;		12963	9291	32665 ATUNTAQUI		3/1/2022	49586	13	2/1/2017	31/12/2021
25087	2022-01-03 11:15:43.7	0002-DPI-AE-2022	Los objetivos: a los gastos;		12963	9291	31677 IBARRA		3/1/2022	49586	13	2/1/2017	31/12/2021
25088	2022-01-03 11:16:40.0	0001-DNA1-C.E.E.-AI-2022	Objetivos: • al origen, reg		27358	0	27358 QUITO		3/1/2022	10/11/2038	13	1/1/2017	31/12/2021
25089	2022-01-03 11:17:57.0	0003-DPI-AE-2022	Los objetivos: a los gastos;		12963	9291	32210 OTAVALO		3/1/2022	49586	13	2/1/2017	31/12/2021
25090	2022-01-03 11:19:41.0	0004-DPI-AE-2022	Los objetivos: a los gastos;		12963	9291	31017 COTACACHI		3/1/2022	49586	13	2/1/2017	31/12/2021
25091	2022-01-03 11:20:35.7	0001-DNA5-GAD-EMSEGUR	Los objetivos: a las fases pr		29515	0	29515 QUITO		3/1/2022	14/2/2024	13	1/1/2017	31/12/2020
25092	2022-01-03 11:21:14.8	0005-DPI-AE-2022	Los objetivos: a las operaci		12963	9291	27955 CAHUASQUI		3/1/2022	49586	13	2/1/2017	31/12/2021
25093	2022-01-03 11:23:00.0	0006-DPI-AE-2022	Los objetivos: a las operaci		12963	9291	27961 PATAQUI		3/1/2022	49586	13	2/1/2017	31/12/2021
25094	2022-01-03 11:27:20.8	0007-DPI-AE-2022	Los objetivos: a los ingreso		12963	9291	32568 SAN JOSE DE		3/1/2022	4/10/2035	13	2/1/2017	31/12/2021
25095	2022-01-03 11:29:42.6	0001-DNA5-GAD-EPMHV-A	Los objetivos: a los ingreso		29429	0	29429 Empresa Púk		3/1/2022	47833	13	2/1/2018	31/12/2021
25096	2022-01-03 11:30:19.7	0001-DNA8-TVIPYA-DGAC-	Los objetivos: a las fases pr		11638	0	11638 QUITO		3/1/2022	50393	13	1/5/2019	31/7/2021
25097	2022-01-03 11:33:54.7	0008-DPI-AE-2022	Los objetivos: a los ingreso		12963	9291	34102 IBARRA		3/1/2022	4/10/2035	13	2/1/2017	31/12/2021
25098	2022-01-03 11:42:06.5	0001-DNA1-CNE-AI-2022	Objetivos: - a los procesc		16384	0	16384 QUITO		3/1/2022	47271	13	18/8/2018	31/12/2021
25099	2022-01-03 11:48:36.0	0001-DNA2-MIDUVI-AI-202	Los objetivos: a los procesc		11925	0	11925 QUITO		3/1/2022	50394	13	1/1/2017	31/12/2021
25100	2022-01-03 11:58:44.0	0001-DPI-GADMI-AI-2022	Los objetivos: al cumplimie		27426	0	27426 Ibarra		3/1/2022	16/3/2036	13	1/6/2017	31/12/2021

Los campos considerados para el estudio se conforman en un archivo en formato Excel con el nombre “SubconjuntoDatos.xlsx”, en este archivo constan los nombres de los campos comprensibles y se detallan en la siguiente tabla.

Tabla 6

Campos Usados para el Estudio

Campo	Descripción
Accion de Control	Número de la acción de control, creada en la planificación
Nombre Entidad AI	Nombre de la entidad donde ejecutó el examen la AI
Tipo Plan	El tipo de plan, puede ser de auditoría externa AE o de auditoría interna AI
Nombre Unidad CGE	Nombre de la unidad que ejecuta el examen, unidad AE de la CGE
Cod Unidad	Código de la unidad que ejecuta la orden de trabajo (OT)
Unidad Ejecuta OT	Nombre de la unidad que ejecuta la OT
Sigla Unidad	Siglas de la unidad que ejecuta la OT
Ambito	Ámbito que corresponde la OT
Cod OT	Código de la orden de trabajo
Num OT	Número de la orden de trabajo, cuando ejecuta AE, se define: 0001-DPI-AE-2022, cuando ejecuta la AI, se define: 0001-DNA2-INPC-AI-2022
Fecha Reg OT	Fecha de registro de la orden de trabajo, en la que inicia su ejecución del examen.
Objetivos OT	Objetivos de la orden de trabajo
Alcance OT	Alcance de la orden de trabajo.
Provincia	Provincia en la que se ejecuta el examen de auditoría

Periodo a Examinar Desde	Periodo de inicio considerado para el examen de auditoría
Periodo a Examinar Hasta	Fecha fin del periodo considerado para el examen de auditoria
Días Equipo OT	Días asignados para la ejecución de la OT al equipo de trabajo (jefe de equipo)
Días Revision	Días asignados al equipo de revisión de la OT, este es un equipo de control de calidad que valida, previo al envío a aprobación del informe
Total Dias Eje	Días asignados para la ejecución de la OT, cuyo resultado es el borrador del informe de auditoría previo al envío de la validación por el equipo de control de calidad y la aprobación del informe.
Entid Relacionadas	Entidades relacionadas a una entidad principal, por ejemplo, el Ministerio de Gobierno, tiene como entidad relacionada a la Comandancia General de Policía
Cod Tipo Examen	Código del Tipo Examen, para el estudio es el numero 13
Tipo Examen	Tipo Examen, de los diferentes tipos, consideramos para el estudio el Examen Especial
Id Informe	Id del Informe
Num Informe	Número de informe de auditoría
Fecha Aprob Informe	Fecha de aprobación del informe de auditoría
Tipo Informe	Tipo de informe, para el caso de estudio se consideró el informe general (G)
Introduccion Inf	Breve introducción del Informe
Limitación Alcance Inf.	Breve resumen de la limitación al alcance del informe
Año Dcto Interno	Año que corresponde al informe de auditoría
Fecha Reg Informe	Fecha del registro del informe en el sistema
Unidad Ejecuta Informe	Unidad que ejecuta el informe
Alcance Informe	Breve resumen del alcance del informe
Seguimiento 2022	Seguimiento de las recomendaciones de auditoría antes del 2022
Dias Ejecucion	Campo generado con el número de días hábiles desde la ejecución, considerando la fecha de inicio de la orden de trabajo hasta la aprobación del informe.
EjecutaTramite?	Campo generado para identificar si se ejecutó la planificación (Ejecutado =1 y No Ejecutado=0).

Clasificación y Selección de Datos para estudio

Para seleccionar los datos de análisis para el presente estudio, se realizó la coordinación y gestión con los Administradores Funcionales de la Dirección de Planificación Institucional de las herramientas informáticas de los procesos de la planificación de acciones de

auditoría, de la ejecución de las acciones de control y aprobación de informes para identificar y seleccionar los campos de cada uno de los procesos utilizados.

Se solicitó al señor Contralor General del Estado, que autorice el uso de datos de los procesos antes mencionado con fines académicos, quien mediante sumilla inserta en Memorando No. 001-CG-2022, solicita a la Dirección de Planificación, misma que tiene a su cargo la administración funcional de los procesos seleccionados para el estudio, realice el análisis pertinente sobre la información solicitada, esta Dirección mediante Memorando No. 0336-DNPyEI-PSyEI-2023 remite datos respecto a la planificación del control y su respectiva ejecución y aprobación de informes del año 2022.

A los datos entregados, se analizan y se seleccionan los datos que corresponden a los procesos considerados para el presente estudio, que son el de planificación, de ejecución de la orden de trabajo y de aprobación del informe de auditoría.

A continuación, se describe el proceso realizado para identificar los campos, los datos y seleccionar la muestra.

Con los datos obtenidos se realizó varios análisis y se fue trabajando y consolidando en consultas más elaboradas que integren y muestren los datos de los tres procesos, llegando a consolidar en una consulta "consulta_ejecucionOT_InformesAprobados.txt" que se lo menciona y consta en un archivo anexo por razones de seguridad y confidencialidad de la información.

Se observa que existen campos que registran datos NULL en el proceso inicial que es el de la planificación, es decir que, en el transcurso del registro de datos no se registra la información completa, por ejemplo, para auditorías externas no registran el dato de la unidad que ejecutará el examen.

De la información suministrada por el Administrador Funcional del Sistema de Control (ejecución de acciones de control), todos los procesos que se ejecutan están integrados, sin embargo, se observa que en la planificación al tener datos NULL o datos vacíos se podría tener

inconvenientes en la ejecución del siguiente proceso que es la orden de trabajo y por ende del informe de auditoría.

Para el estudio, se seleccionan unidades estratégicas, ámbitos de control, tipo de examen con ejecución de acciones de control, a continuación, se describen los parámetros y campos seleccionados:

Unidades y Ámbitos:

DNA1 Dirección Nacional de Auditoría de Administración Central

- Administrativo

DNA2 Dirección Nacional de Auditoría de Sectores Sociales

- Bienestar Social
- Educación

DNA4 Dirección Nacional de Auditoría de Telecomunicaciones, conectividad y sectores productivos

- Telecomunicaciones

DNA6 Dirección Nacional de Auditoría de Recursos Naturales

- Ambiente
- Recursos naturales

Año: 2022

Tipos de exámenes:

- Exámenes Especiales

Tipos de Informes:

- Generales

Indicadores de evaluación:

- Número de exámenes planificados
- Número de exámenes cumplidos

- Número de informes aprobados
- Tiempo de ejecución de la orden de trabajo de una acción de control.

Preprocesamiento de datos

Los datos obtenidos y depositados en formato xlsx, se los analiza campo por campo y se va seleccionando los que aportan valor según los objetivos planteados, con la ejecución del preprocesamiento de datos, se filtran aquellos que se requieren ser eliminados como los valores nulos, incorrectos, no válidos, desconocidos, según las necesidades de los algoritmos que están siendo usados.

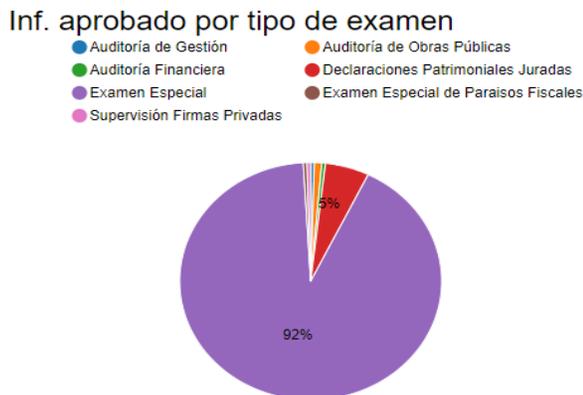
A los datos del archivo digital, se realizó un estudio y análisis y se consideran retirar los tipos de exámenes de operativo de control vehicular e imprevistos, por cuanto estos registros no conllevan un informe aprobado, se retiran también los cancelados del estado del plan, y se seleccionan los campos requeridos para el estudio.

Para la clasificación y preprocesamiento de los datos se utilizó la herramienta Knime para así tener una previsualización de datos mediante la construcción de un primer algoritmo de árbol de decisión en la ruta y con el nombre “knime://LOCAL/LecturaDatos/PlanificacionAprobada_InfReq_2022”.

Luego de la previsualización de datos mediante el nodo “Pie/Donut Chart” se toman los tipos de examen “Examen Especial” que son los exámenes de mayor porcentaje, que se realizan frecuentemente y que terminan con la aprobación del informe, como se observa en la Figura 12.

Figura 12

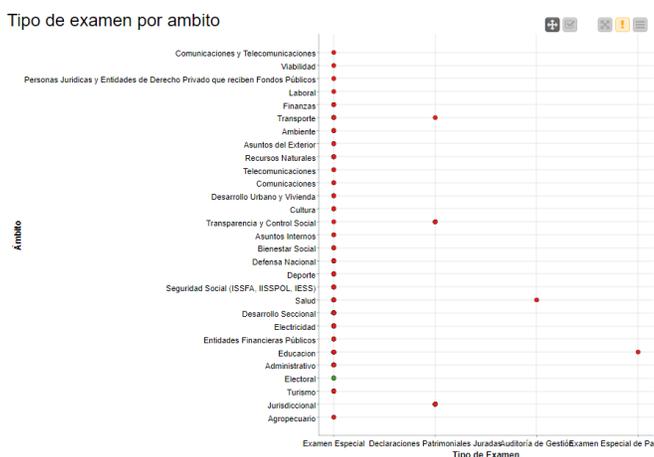
Informes Aprobados por Tipo de Examen.



Así mismo mediante el nodo de visualización “scatter plot” de knime se observa que el examen especial se ha ejecutado para todos los ámbitos, esta característica, se visualiza en la Figura 13.

Figura 13

Tipos de Examen por Ámbito.



Transformación

Esta etapa se lo realizó utilizando la herramienta Knime, se inició con la generación de un archivo procesado que se utilizará para la minería de datos. Para ello se identifica los campos que requieren ser transformados, dentro de lo cual se trabajan en los campos de fecha, a través de la herramienta SQL Server, se procede a dar el formato de fecha, se generó un campo para identificar los días ejecutados desde el inicio de la orden de trabajo hasta la

aprobación del informe de auditoría, mismo que servirá para el análisis del tiempo de ejecución del informe.

Dentro del archivo nuevo generado en formato xlsx, denominado SubconjuntoDatosVs01.xlsx se procede a reemplazar algunos datos de los campos que tienen el valor NULL con el valor blanco, " ".

Así mismo se agregó un campo necesario para el estudio y ejecución del algoritmo designado con el nombre "Examen Cumplido", conformándolo con el valor "0" si no está cumplido y con el valor "1" si hasta la fecha en la que se obtuvo el archivo se encuentra cumplido.

Minería de datos

En esta etapa, se seleccionó las técnicas de minería de datos adecuadas para cumplir con los objetivos, se hacen uso de las funciones y algoritmos que posee la herramienta Knime y Weka y se construyen los algoritmos con los siguientes modelos de minería de datos: Árboles de decisión, Naive Bayes y K-Means.

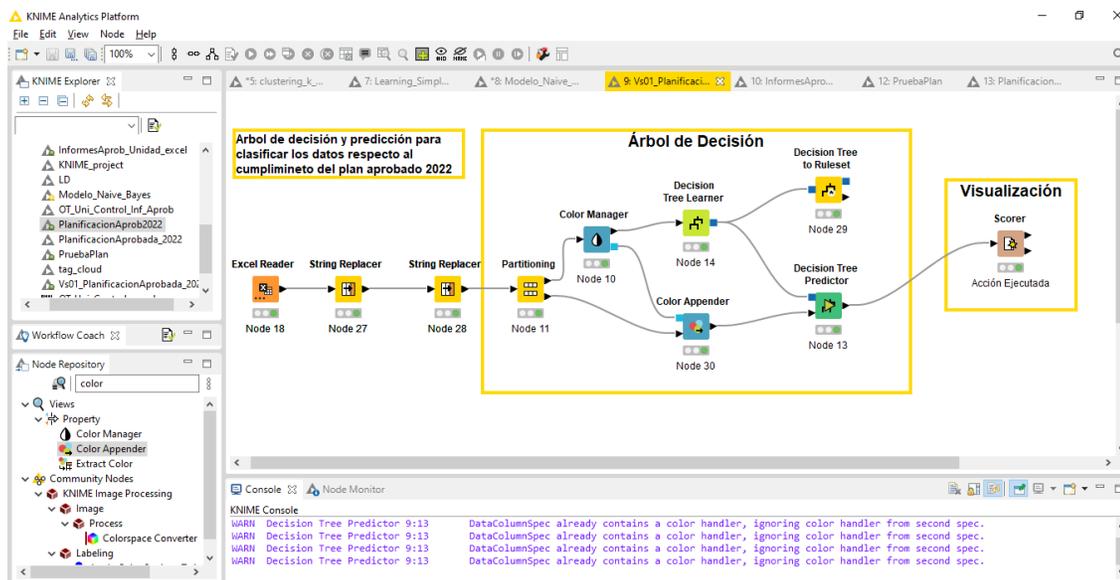
Árboles de decisión

La construcción del modelo predictivo con el algoritmos de árboles de decisión nos permite identificar las variables relevantes o con mayor nivel de correlación y que son las que intervienen en la predicción si una acción de control planificada en el año 2022 tiene emitido un informe aprobado.

En la Figura 14, se representa el modelo de árboles de decisión construido en la herramienta Knime.

Figura 14

Modelo de Clasificación de Datos con árboles de decisión en Knime

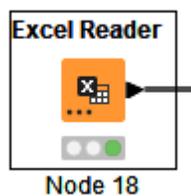


Para la implementación del modelo de minería de datos del algoritmo de Árboles de Decisión y Predicción, se utilizó los nodos que se describen a continuación desde la Figura 15 hasta la Figura 21:

Excel Reader

Figura 15

Nodo Excel Reader

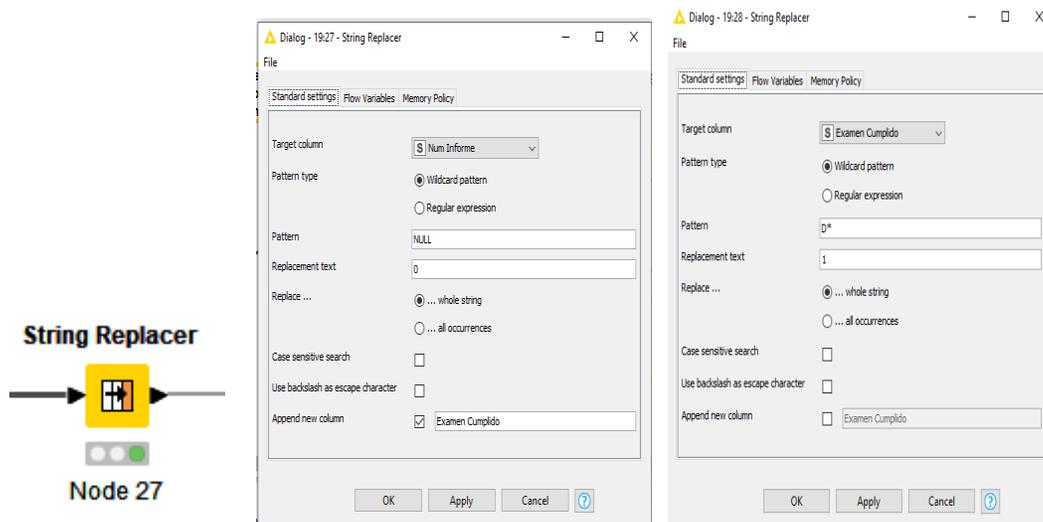


Nodo usado para abrir y leer los archivos en formato xlsx, este nodo permite obtener todos los campos del archivo que forman parte de la selección de datos para el estudio.

String Replacer

Figura 16

Nodo String Replacer y su Configuración

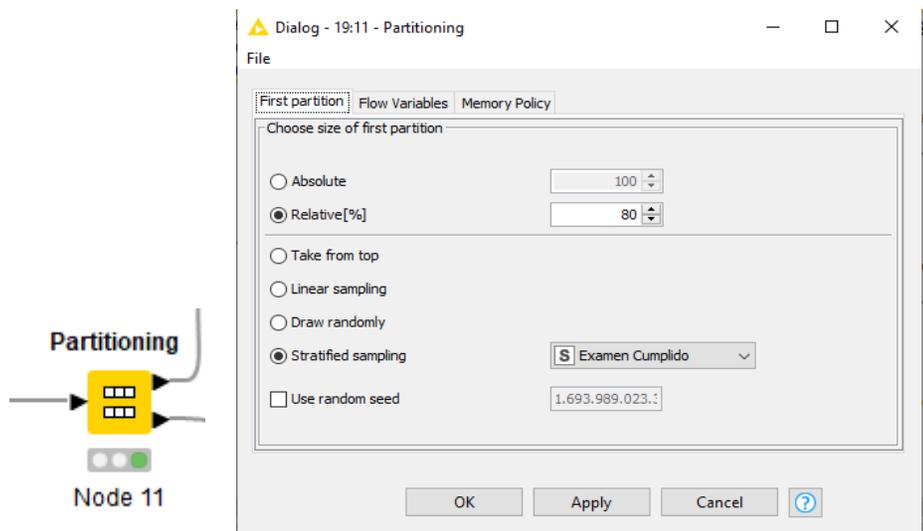


El Nodo String Replacer, permite reemplazar valores en campos tipo string, según cumplan las condiciones con un determinado patrón de comodines.

Partitioning

Figura 17

Nodo Partitioning

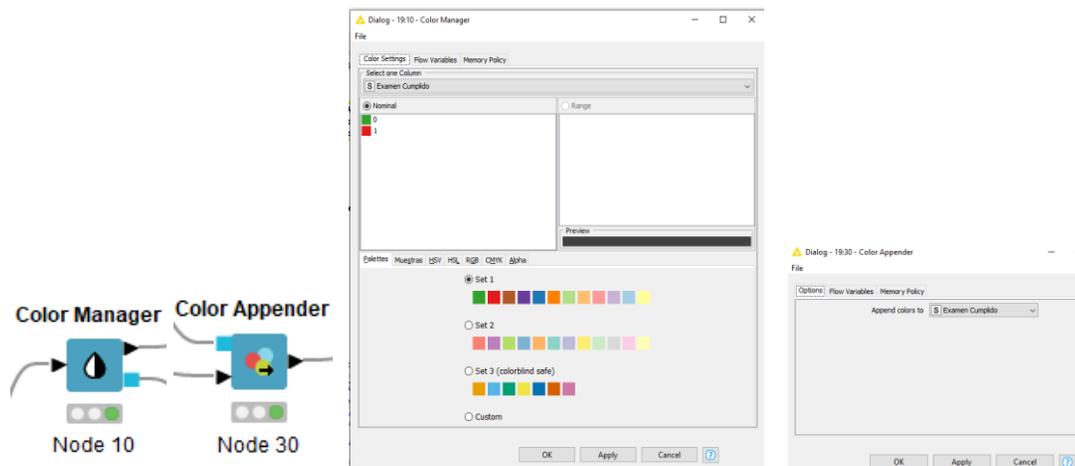


El nodo llamado "Partitioning" realiza la partición de la data en una relación 80% para el aprendizaje del modelo y el 20% restante para las pruebas que servirá para la ejecución según el modelo construido.

Color Manager y Color Appender

Figura 18

Nodo Color Manager y Color Appender y su Configuración



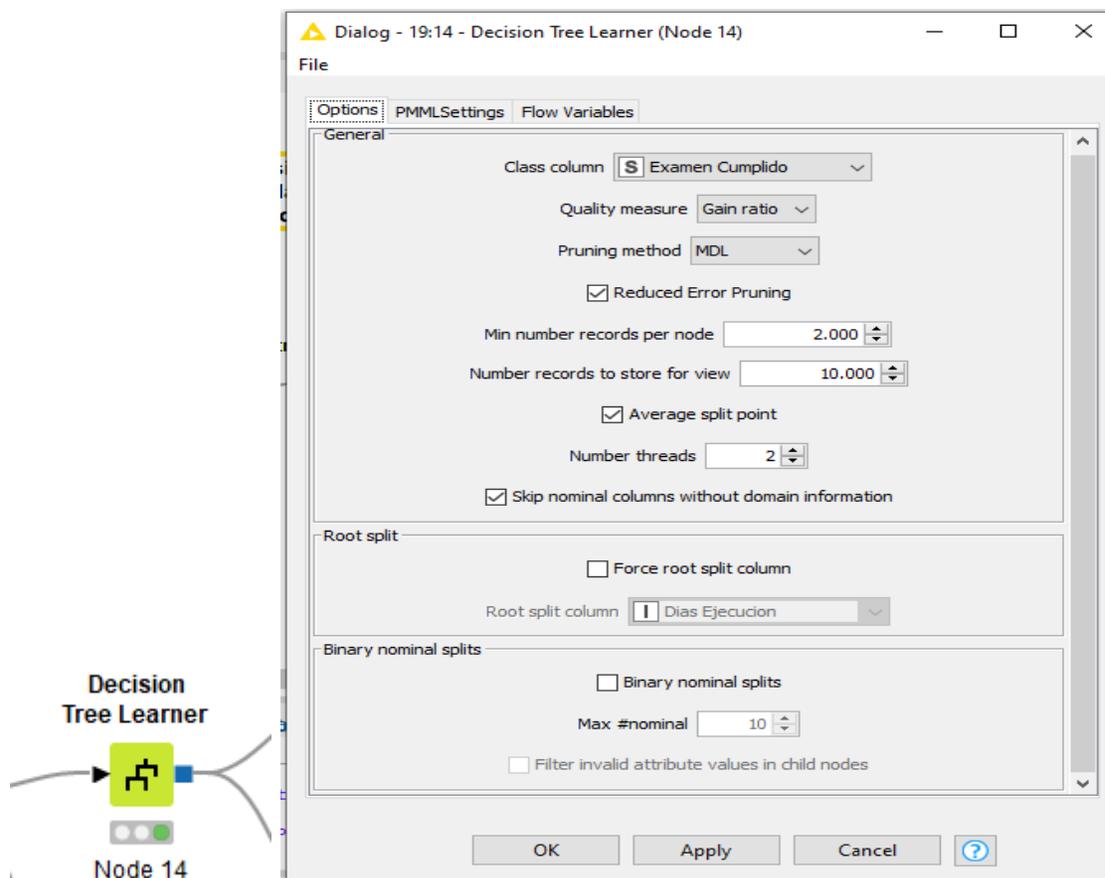
El nodo Color Manager, permite asignar un color específico a los valores de un determinado campo tipo string, con la finalidad de mejorar la interpretación de los resultados al aplicar el modelo a los datos del archivo de la entrada. Para el estudio se configuró para diferenciar las acciones que se han ejecutado.

El nodo Color Appender es usado para agregar los colores a los campos de la tabla de las columnas configuradas en el nodo Color Manager

Decision Tree Learner

Figura 19

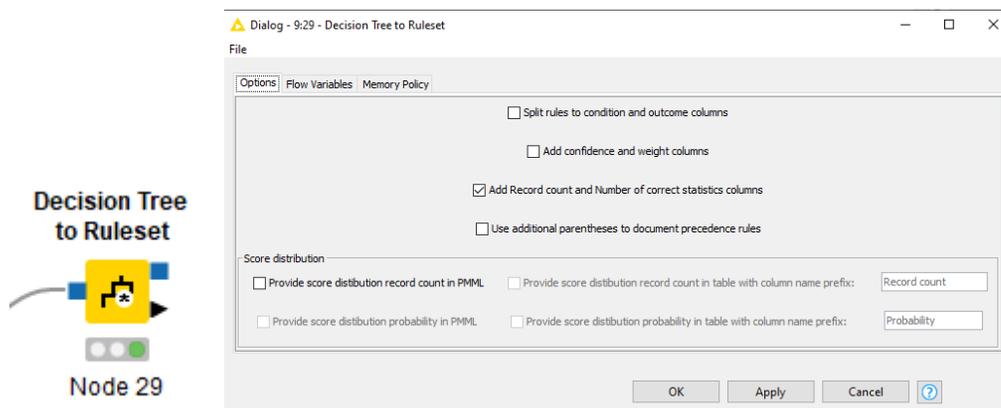
Nodo Decision Tree Learner y su Configuración



El nodo “Decisión Tree Learner”, es usado para el aprendizaje del modelo del árbol de decisión, para el estudio se configuró en la sección clase a la columna “Examen Cumplido”, en la medida de calidad, se configuró “Gain ratio”, en el método de poda se seleccionó “MDL”, se chequeó la opción de “Reduce Error Pruning” en el número mínimo de registro por nodo se configuró 2,000, en el número de registros a almacenar para la vista se configuró 10,000.

Figura 20

Nodo Decision Tree Rulest y su Configuración

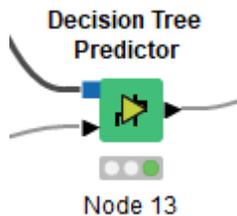


Nodo usado para visualizar las reglas resultantes de forma textual, mismas que se aplican en la ejecución del modelo construido.

Decision Tree Predictor

Figura 21

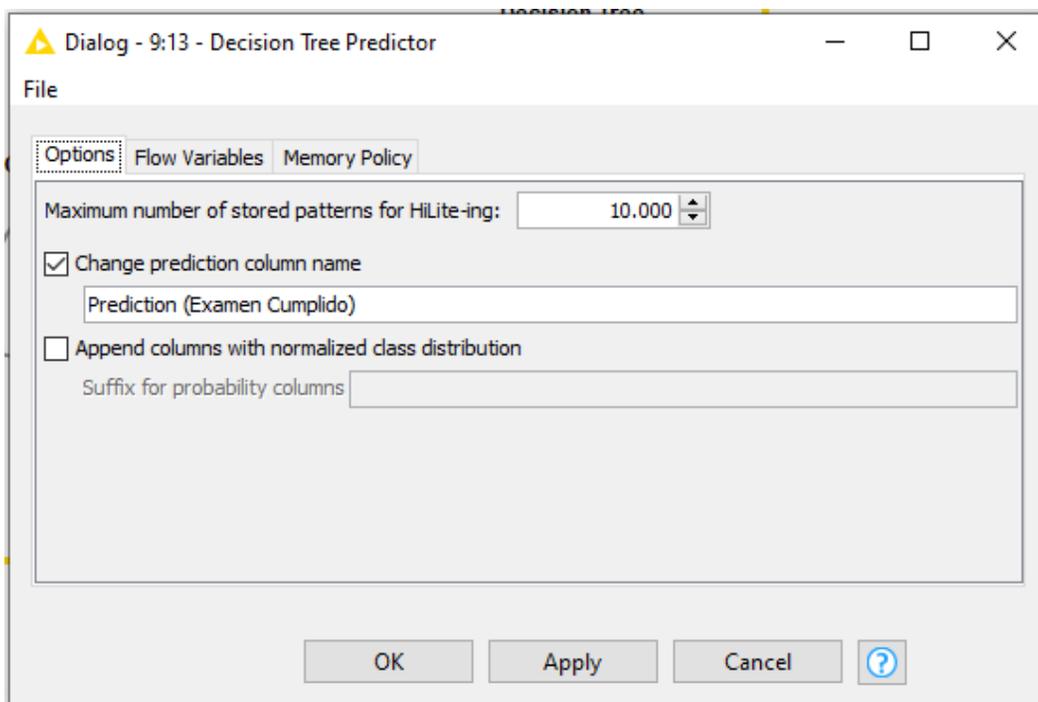
Nodo Decision Tree Predictor



El nodo Decision Tree Predictor, se usa para pronosticar la data de prueba en base al entrenamiento brindado al nodo “Decision Tree Learner”, para el caso de estudio nos permite predecir el valor de la clase “Examen Cumplido” para nuevos patrones que se muestra en la sección “Prediction (Examen Cumplido)”. En la Figura 22 se presenta la configuración del nodo.

Figura 22

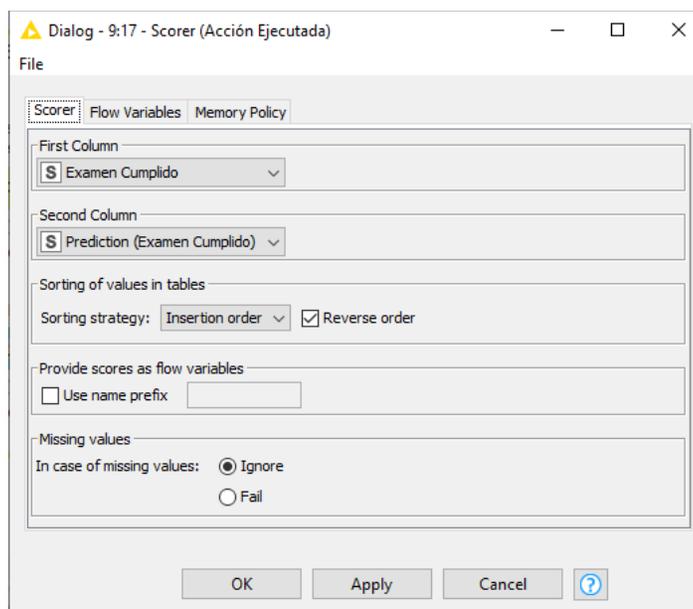
Representación de la Configuración del Nodo Decision Tree Predictor



Scorer

Figura 23

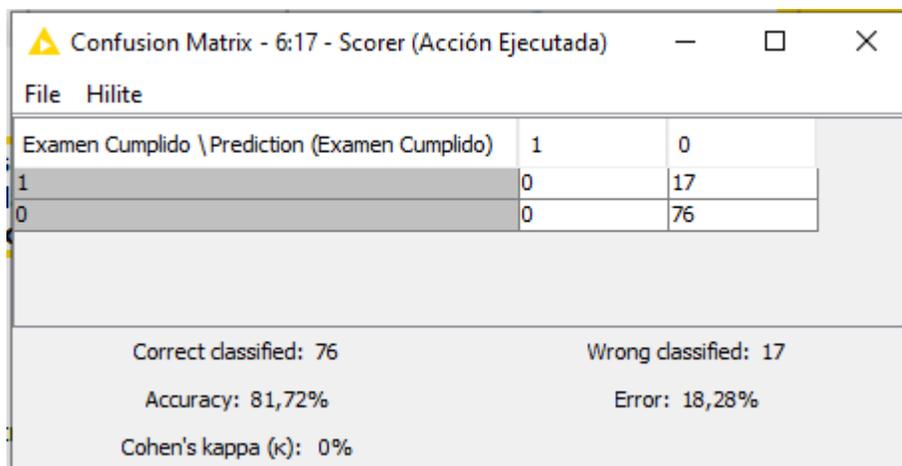
Nodo Scorer y su configuración



El nodo Scorer, se utiliza para la visualización de la salida de los resultados en la “Matriz de Confusión” se cuentan el número de coincidencias en cada celda, al comparar el campo de entrenamiento con el campo predicho, este resultado permite evaluar que tan bueno resulta ser el modelo aplicado con las variables seleccionadas. Para el estudio se muestra si una acción de control se ejecuta.

Figura 24

Resultado de la Matrix Confusión (Nodo Scorer) de Árboles de Decisión



Examen Cumplido \ Prediction (Examen Cumplido)	1	0
1	1	17
0	0	76

Correct classified: 76 Wrong classified: 17
Accuracy: 81,72% Error: 18,28%
Cohen's kappa (κ): 0%

Decision Tree Learner con diferente Configuración

Figura 25

Nodo Decision Tree Learner y su Configuración diferente

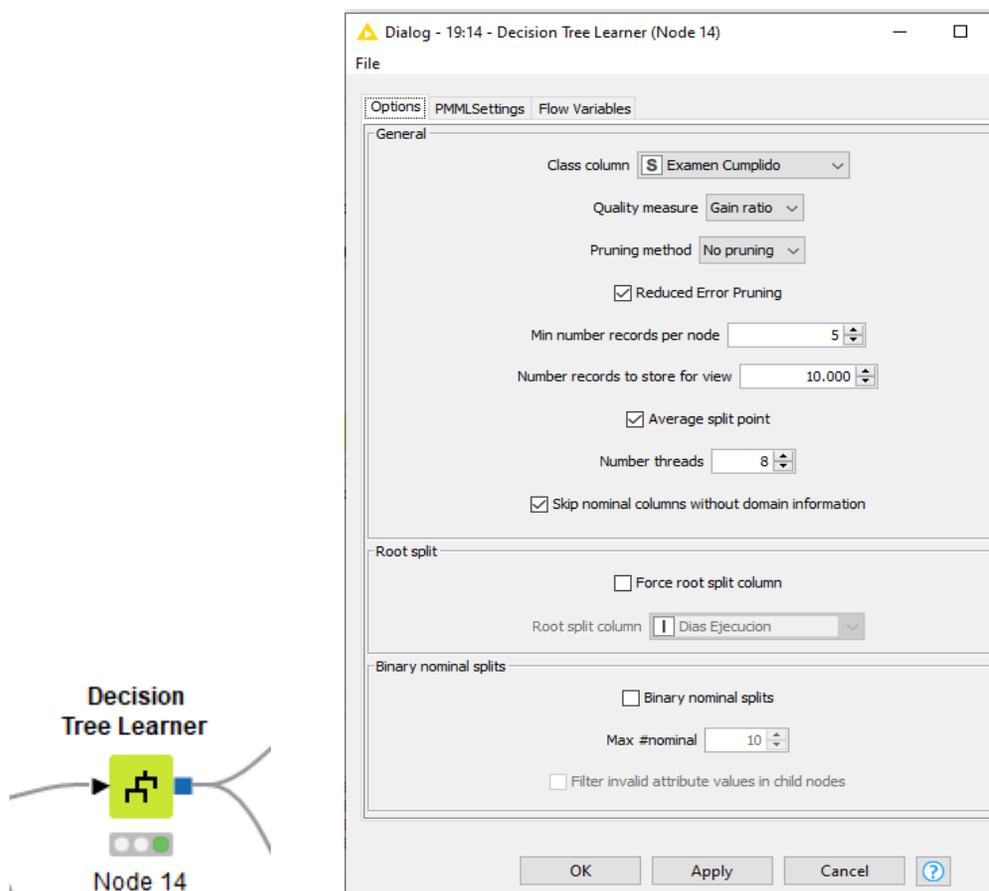


Figura 26

Resultado de la Matrix Confusión (Nodo Scorer) de Árboles de Decisión

Confusion Matrix - 19:17 - Scorer (Acción Ejecutada)		
Examen C...	1	0
1	17	0
0	0	76

Correct classified: 93 Wrong classified: 0
 Accuracy: 100% Error: 0%
 Cohen's kappa (κ): 1%

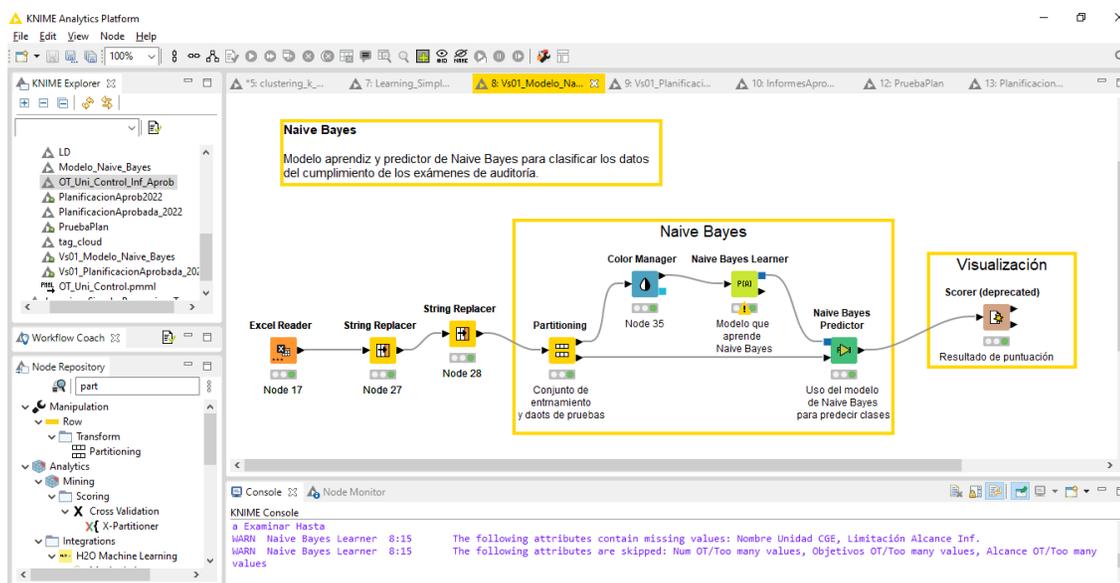
Naive Bayes

La aplicación de la técnica de minería de datos Naive Bayes, nos ayuda a identificar las variables relevantes o con mayor grado de correlación, que intervienen en el cumplimiento de la acción de control, se define con la variable “Examen Cumplido”.

En la Figura 27 se representa el algoritmo del Modelo Naive Bayes

Figura 27

Modelo de Naive Bayes

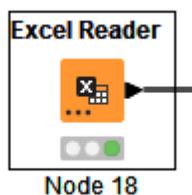


Para la implementación del modelo de minería de datos del algoritmo de predicción de Naive Bayes, se utilizaron los nodos que se describen a continuación desde la Figura 28 hasta la Figura 39:

Excel Reader

Figura 28

Nodo Excel Reader

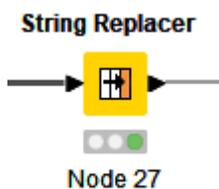


Nodo usado para abrir y leer archivos en formato xlsx, este nodo permite obtener todos los campos del archivo que forman parte de la selección de datos para el estudio.

String Replacer

Figura 29

Nodo String Replacer

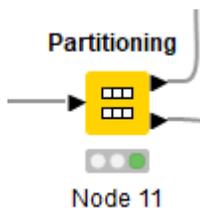


El Nodo String Replacer, permite reemplazar valores en campos tipo string, según cumplan las condiciones con un determinado patrón de comodines.

Partitioning

Figura 30

Nodo Partitioning

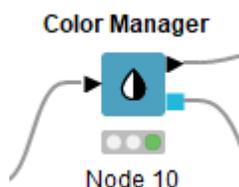


El nodo llamado "Partitioning" realiza la partición de la información en una relación 70% para el aprendizaje del modelo y el 30% restante para las pruebas que servirá para la ejecución según el modelo construido.

Color Manager

Figura 31

Nodo Color Manager

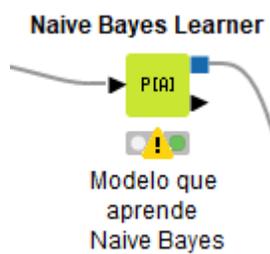


El nodo Color Manager, permite asignar un color específico a los valores de un determinado campo tipo string, con la finalidad de mejorar la interpretación de los resultados al aplicar el modelo a los datos del archivo de la entrada. Para el estudio se configuró para diferenciar las acciones que se han ejecutado.

Naive Bayes Learner

Figura 32

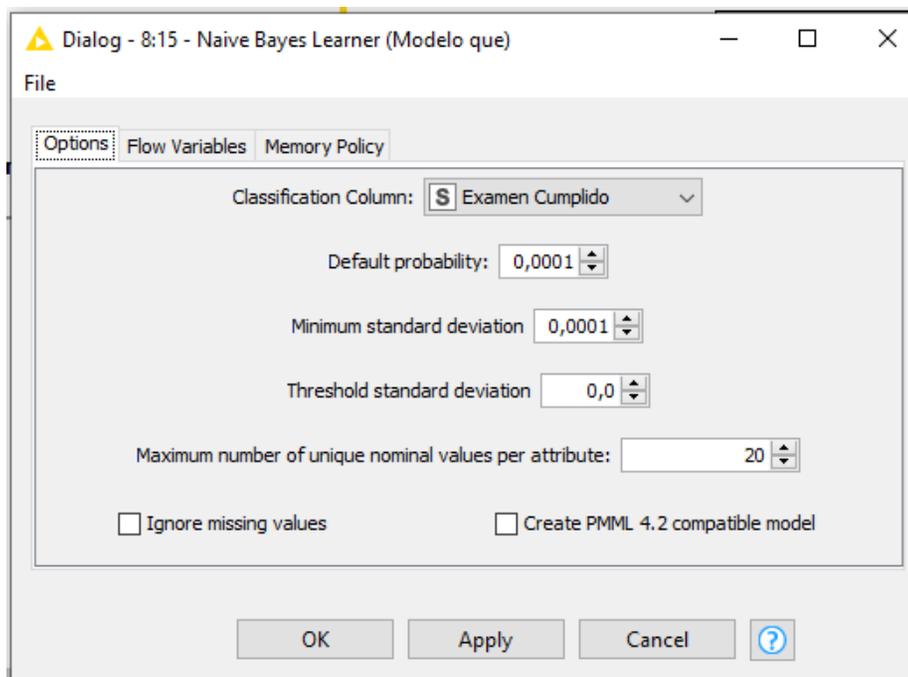
Nodo Naive Bayes Learner



Este nodo es usado para el aprendizaje de la técnica Naive Bayes, para el estudio se configuró en la sección de classification column la variable “Examen Cumplido”, en Default probability se configuró el valor 0,0001, entre otras configuraciones. En la Figura 24 se muestra las confirmaciones realizadas.

Figura 33

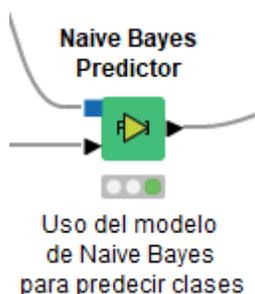
Configuraciones del Nodo Naive Bayes Learner



Naive Bayes Predictor

Figura 34

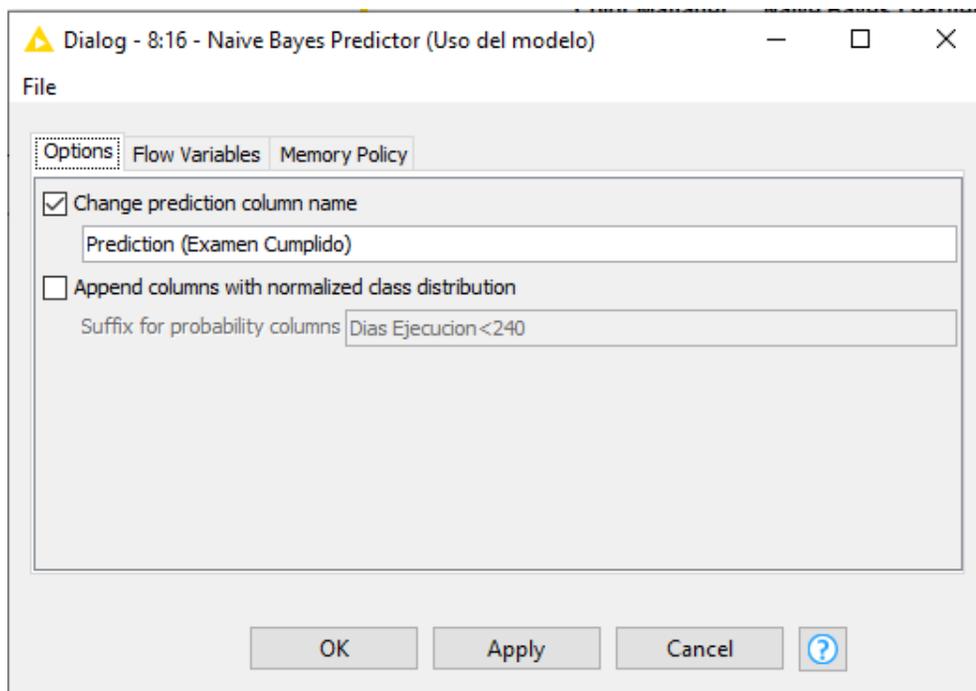
Nodo Naive Bayes Predictor



Este nodo es usado para predecir la data de prueba en base al entrenamiento dado en el nodo “Naive Bayes Learner”, para el presente estudio este nodo nos permite predecir el valor de la columna “Examen Cumplido” para nuevos patrones. En la Figura 37 se muestran las configuraciones del nodo.

Figura 35

Configuraciones del Nodo Naive Bayes Predictor



Scorer

Figura 36

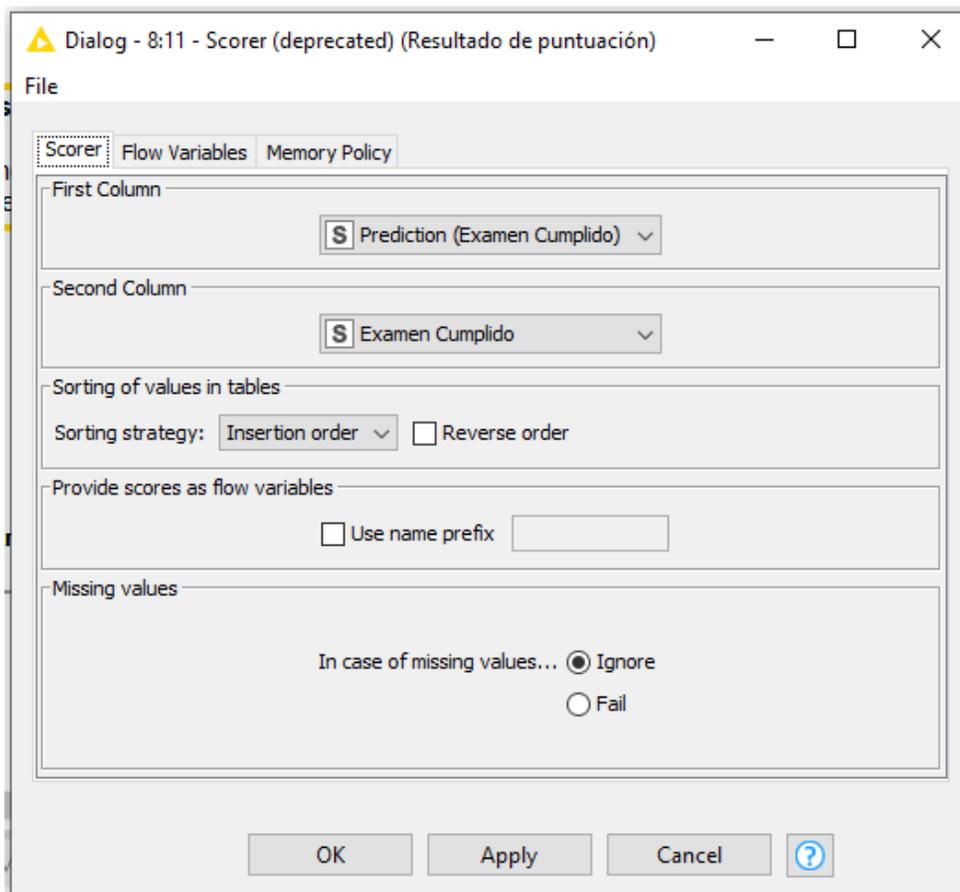
Nodo Scorer



El nodo Scorer, nos permite observar los resultados que presenta la matriz de confusión, se cuentan el número de coincidencias en cada celda, al comparar el campo de entrenamiento con el campo predicho, resultado que permite evaluar que tan bueno es el modelo. Para el estudio mostrará el resultado al evaluar los registros del campo “Examen Cumplido” con el campo Prediction(Examen Cumplido), con el fin de interpretar la confiabilidad y precisión de la construcción del modelo de Naive Bayes. A continuación, se muestra la configuración del nodo Scorer.

Figura 37

Configuración del Nodo Scorer



En la Figura 38, se presenta los resultados a través de la matriz de confusión (nodo scorer)

Figura 38

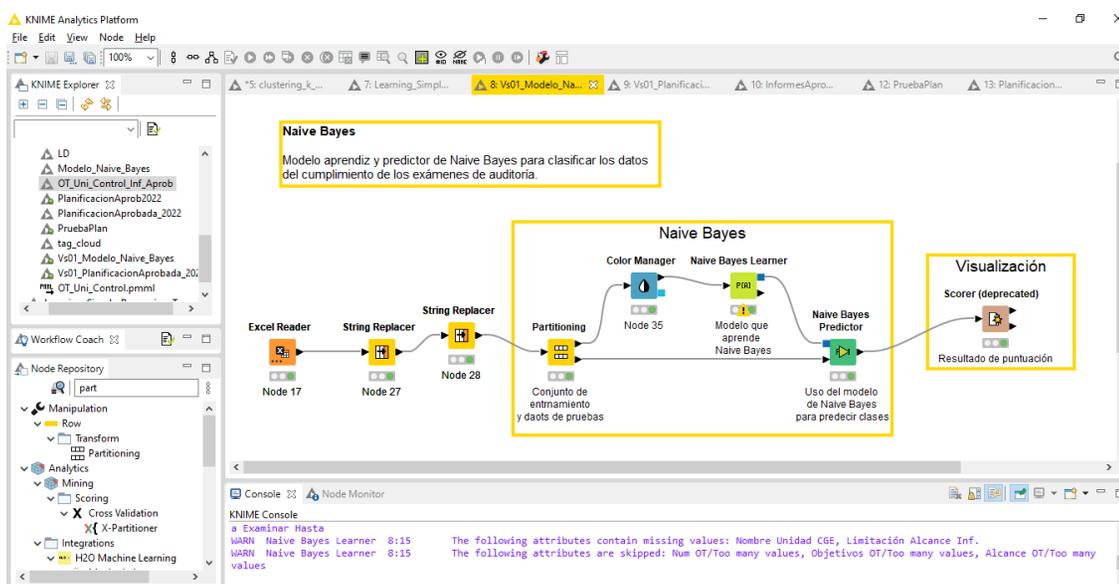
Resultado de la Matriz de Confusión (Nodo Scorer) de Naive Bayes

Confusion Matrix - 4:11 - Scorer (deprecated) (Resul...		
File Hilite		
Prediction (Examen Cumplido) \ Examen Cumplido	0	1
0	76	0
1	0	17

Correct classified: 93 Wrong classified: 0
 Accuracy: 100% Error: 0%
 Cohen's kappa (κ): 1%

Figura 39

Modelo de Naive Bayes



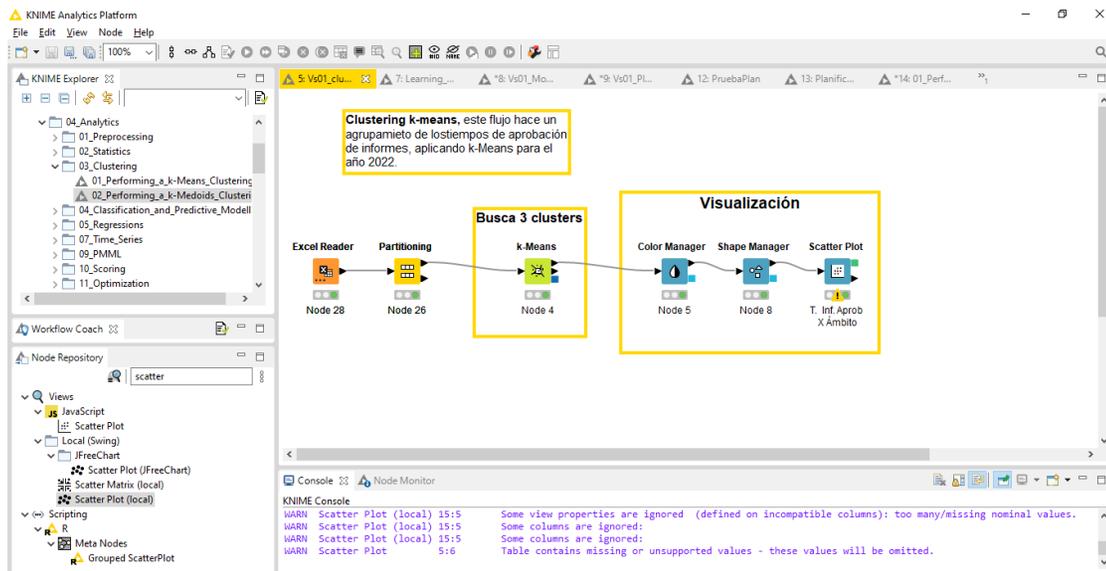
Modelo K-Means

El proceso utilizado en este análisis es la aprobación de informes con el periodo de tiempo, este modelo hace un agrupamiento de los tiempos de aprobación, aplicando la técnica de k-Means para el conjunto de datos seleccionado, se aplica a la variable "Dias Ejecucion" generada con los periodos de tiempo que transcurre desde la fecha de inicio de la ejecución de la orden de trabajo hasta la aprobación del informe.

En la Figura 40, se muestra la construcción del modelo de K-Means

Figura 40

Modelo K-Means

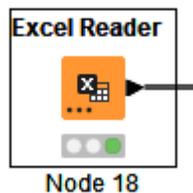


Para la implementación del modelo de minería de datos con el algoritmo de clustering de K-Means, se utilizaron los nodos que se describen a continuación desde la Figura 41 hasta la Figura 49:

Excel Reader

Figura 41

Nodo Excel Reader

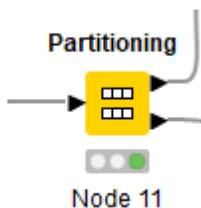


Nodo usado para abrir y leer archivos en formato xlsx, este nodo permite obtener todos los campos del archivo que forman parte de la selección de datos para el estudio.

Partitioning

Figura 42

Nodo Partitioning

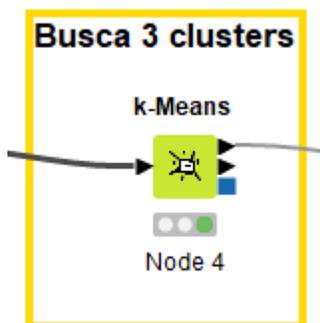


El nodo llamado “Partitioning” realiza la partición de la información en una relación 80% para el aprendizaje del modelo y el 20% restante para las pruebas que servirá para la ejecución según el modelo construido.

K-Means

Figura 43

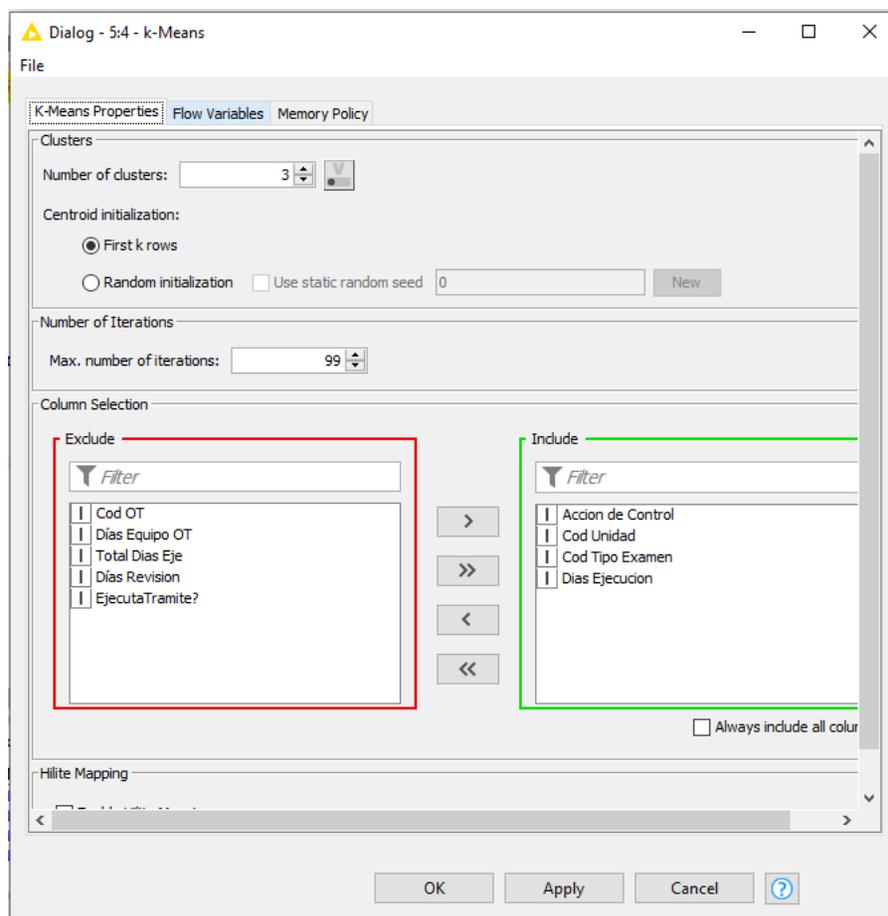
Nodo K-Means



Este nodo es usado para el aprendizaje de la técnica de clustering con K-Means, para el estudio se configuró en la sección de Number of clusters 3 clusters para las agrupaciones, las demás configuraciones se dejó las que son por defecto, en centroids initialization, seleccionado First k rows, y en el número de iteraciones seleccionado 99, entre otras configuraciones que se muestran en la Figura 44.

Figura 44

Configuración del Nodo K-Means

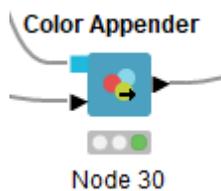


Para la visualización tenemos los siguientes nodos:

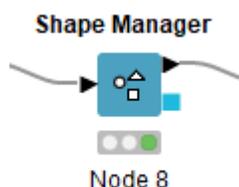
Color Manager

Figura 45

Nodo Color Manager



Este nodo es usado para agregar los colores a los campos de la tabla de las columnas configuradas en el nodo Color Manager

Figura 46*Nodo Shape Manager*

El nodo Shape Manager, asigna formas (diferentes) para cada valor de atributo de una columna, es decir, para cada valor posible, con la finalidad de mejorar la interpretación de los resultados al aplicar el modelo a los datos del archivo de la entrada.

Scatter Plot

Figura 47*Nodo Scatter Plot*

Este nodo es un gráfico de dispersión usado para la visualización de los resultados al aplicar el algoritmo de clustering k-means, que utiliza una biblioteca de gráficos basada en JavaScript.

La configuración del nodo scatter plot, permite elegir el tamaño de una muestra para mostrar y habilitar ciertos controles, que luego están disponibles en la vista, esto es la posibilidad de elegir diferentes columnas para x e y o la posibilidad de establecer un título. En la Figura 51 se muestran las configuraciones del nodo.

Figura 48*Configuración del Nodo Scatter Plot*

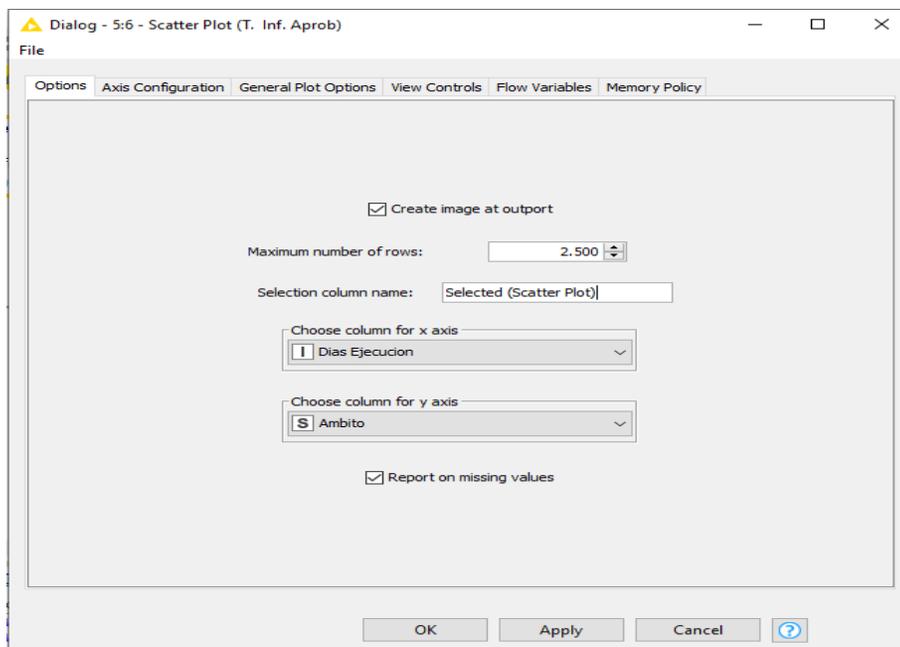


Figura 49

Resultado del nodo Scatter Plot de K-means

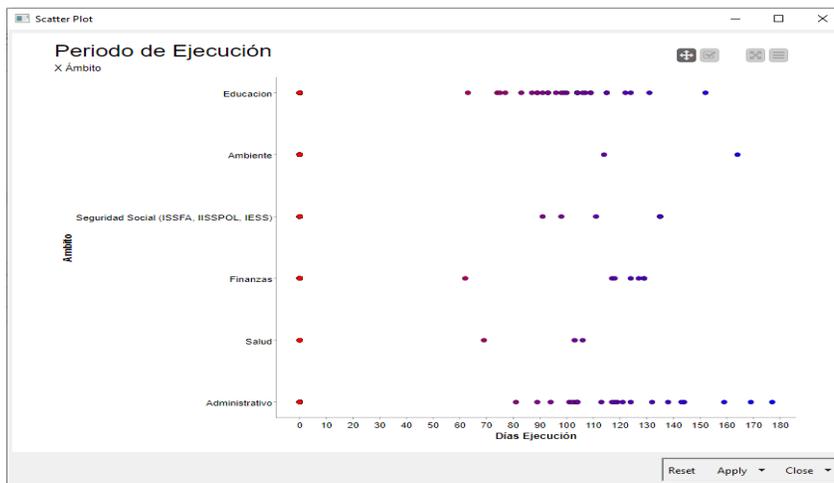
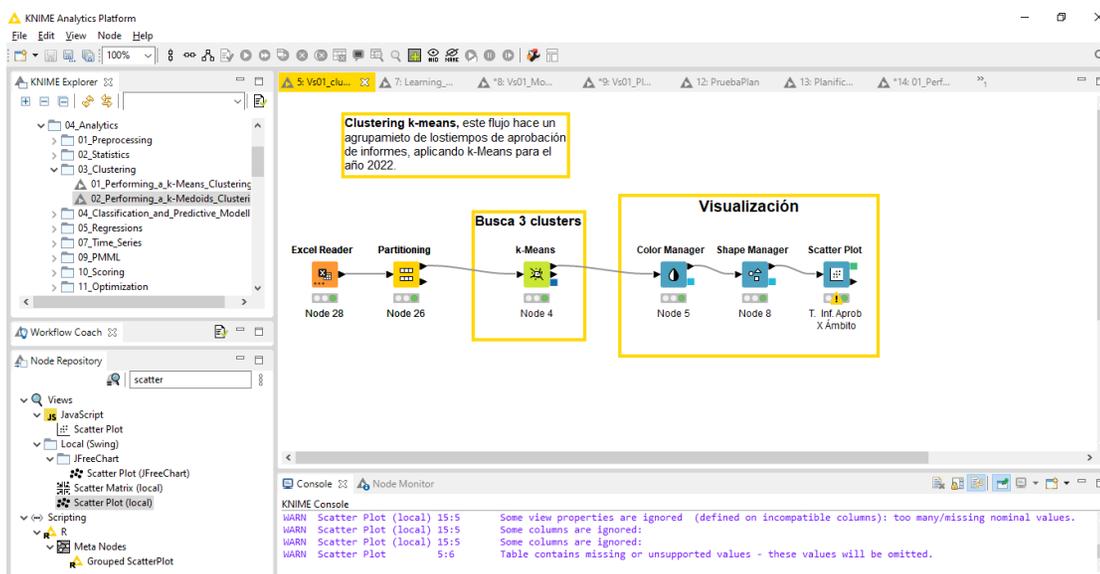


Figura 50

Modelo K-Means



Para el presente estudio se propuso tres modelos, para cada uno de los flujos de trabajo de los modelos desarrollados se incorporaron nodos que permiten la visualización de los resultados, en los 2 modelos propuestos de árboles de decisión y Naive Bayes se incorporó el nodo “Scorer”, con el objetivo de obtener la matriz de confusión de cada modelo y así poder evaluar cada uno de ellos, en el tercer modelo de K-Means se incorporó el nodo Scatter Plot, que permite evaluar los resultados a través de un gráfico de dispersión donde se ubican la mayor concentración de los puntos de un plano cartesiano, mostrando la correlación, pudiendo ser la correlación fuerte, débil o ninguna.

La comparación de los tres modelos permitió evaluar cada uno de ellos y poder determinar cuál de los modelos planteados es el que mejor se adapta a nuestro caso de estudio, el que más aciertos de predicción tenga, así como el menor porcentaje de error, se pudo observar la correlación que existe entre el tipo de examen ejecutado y el tiempo de cumplimiento.

En las Figuras 51 y 52 se observan las matrices de confusión resultantes de cada uno de los modelos.

Figura 51

Matriz de Confusión del Modelo de Árboles de Decisión

Examen C...	1	0
1	0	17
0	0	76

Correct classified: 76 Wrong classified: 17
 Accuracy: 81,72% Error: 18,28%
 Cohen's kappa (κ): 0%

Figura 52

Matriz de Confusión del Modelo de Naive Bayes

Examen Cumplido \ Prediction (Examen Cumplido)	1	0
1	35	0
0	0	26

Correct classified: 61 Wrong classified: 0
 Accuracy: 100% Error: 0%
 Cohen's kappa (κ): 1%

En base a las figuras anteriores, se realiza una comparación entre las matrices de confusión, tal y como se muestra en la siguiente Tabla 7.

Tabla 7

Comparación de Resultados según las Matrices de Confusión generadas en los Modelos de Predicción.

Técnica	Correctamente clasificados	Exactitud	Coficiente Kappa	Clasificados Incorrectamente	Error
Árbol de decisión	76	81,72	0	17	18,28
Naive Bayes	61	100	1	0	0

Análisis de variables con Weka

Para proceder con el análisis de datos en la herramienta Weka se realizó la correspondiente minería de datos y se logró preparar los datos para el correspondiente análisis con la siguiente información: 11 columnas y 73 filas de datos. En la Figura 55, se presenta el conjunto de datos obtenidos.

Figura 53

Conjunto de Datos para el Análisis

A	B	C	D	E	F	G	H	I	J	K	L
Acción de Cor	Nombre Entidad AI	Tipo Plan	Cod Unidad	Unidad Ejecuta OT	Sigla Unidad	Ambito	Cod OT	Num OT	Fecha Reg O	Provincia	Periodo a
27219	SERVICIO DE RENTAS INTERNAS (SRI)	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	25144	0001-DNA3-SRI	03/01/2022		01/01/20
27220	SERVICIO DE RENTAS INTERNAS (SRI)	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	26740	0002-DNA3-SRI	16/05/2022		01/01/20
27221	SERVICIO DE RENTAS INTERNAS (SRI)	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	27885	0003-DNA3-SRI	12/09/2022		01/01/20
27255	MINISTERIO DE ECONOMIA Y FINANZAS	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	25146	0001-DNA3-ME	03/01/2022		01/09/20
27260	MINISTERIO DE ECONOMIA Y FINANZAS	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	26741	0002-DNA3-ME	16/05/2022		01/01/20
27261	MINISTERIO DE ECONOMIA Y FINANZAS	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Finanzas	27854	0003-DNA3-ME	06/09/2022		01/07/20
27343	EMPRESA PÚBLICA METROPOLITANA DE GESTION INTEGR	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Ambiente	25258	0001-DNA5-GA	03/01/2022		01/10/20
27345	EMPRESA PÚBLICA METROPOLITANA DE GESTION INTEGR	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Ambiente	26644	0002-DNA5-GA	09/05/2022		01/01/20
27347	EMPRESA PÚBLICA METROPOLITANA DE GESTION INTEGR	AI	7120	DIRECCIÓN DE AUDITORÍA DE LA PRODUCCIÓN AMBIENTE Y FINAN	DAPAYF	Ambiente	27875	0003-DNA5-GA	09/09/2022		01/05/20
27354	MINISTERIO DE SALUD PÚBLICA	AI	9380	DNA 7 - DIRECCIÓN NACIONAL DE AUDITORIA DE SALUD Y SEGURII	DNA7-SYSS	Salud	25327	0002-DNA7-SyS	04/01/2022		01/01/20
27355	MINISTERIO DE SALUD PÚBLICA	AI	9380	DNA 7 - DIRECCIÓN NACIONAL DE AUDITORIA DE SALUD Y SEGURII	DNA7-SYSS	Salud	25325	0001-DNA7-SyS	04/01/2022		01/01/20
27356	MINISTERIO DE SALUD PÚBLICA	AI	9380	DNA 7 - DIRECCIÓN NACIONAL DE AUDITORIA DE SALUD Y SEGURII	DNA7-SYSS	Salud	27085	0004-DNA7-SyS	06/06/2022		01/01/20
27375	UNIVERSIDAD ESTATAL DE MILAGRO	AI	7320	DELEGACIÓN PROVINCIAL DE GUAYAS		DR1-DPGY	25201	0001-DPGY-UN	03/01/2022	GUAYAS	01/01/20
27376	UNIVERSIDAD ESTATAL DE MILAGRO	AI	7320	DELEGACIÓN PROVINCIAL DE GUAYAS		DR1-DPGY	26590	0002-DPGY-UN	03/05/2022	GUAYAS	01/08/20
27377	UNIVERSIDAD ESTATAL DE MILAGRO	AI	7320	DELEGACIÓN PROVINCIAL DE GUAYAS		DR1-DPGY	27925	0003-DPGY-UN	15/09/2022	GUAYAS	01/01/20
27380	ESCUELA SUPERIOR POLITECNICA DEL LITORAL (ESPOL)	AI	9200	DIRECCIÓN PROVINCIAL DE GUAYAS		DPGY	26633	0002-DPGY-ESF	04/05/2022	GUAYAS	01/01/20

Campos obtenidos para el Estudio

Una vez que se dispone de los datos, se procede a cargar la información en la herramienta de análisis WEKA. En la Figura 55, se presenta la interface de inicio de la herramienta y en la Figura 56 la interface de carga de datos y análisis de información.

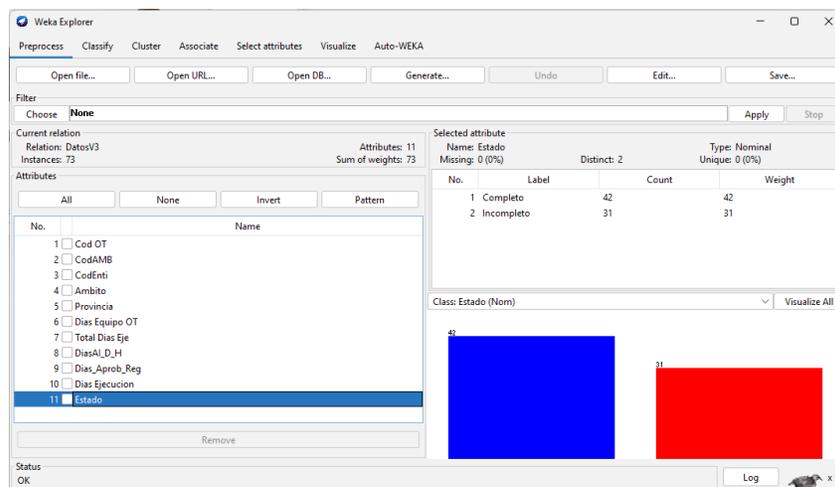
Figura 54

Interface de la Herramienta Weka



Figura 55

Interface de carga de datos y análisis



Análisis de los campos

Una vez que se cargan los datos se puede obtener el resultado de cada uno de los atributos

Figura 56

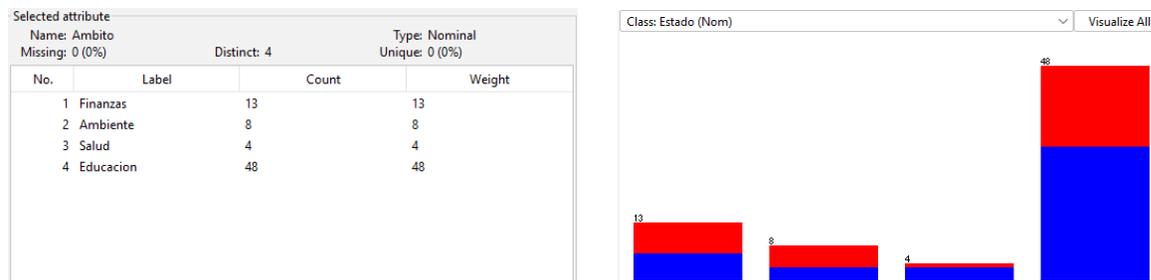
Visualización con Unidad que ejecuta la Orden de Trabajo



En el análisis se encontraron 11 Unidades que ejecutan un número de órdenes de trabajo.

Figura 57

Ejecución de la Orden de Trabajo según el ámbito



Se indica el número de trabajos en los 4 ámbitos (Finanzas, Ambiente, Salud y Educación)

Figura 58

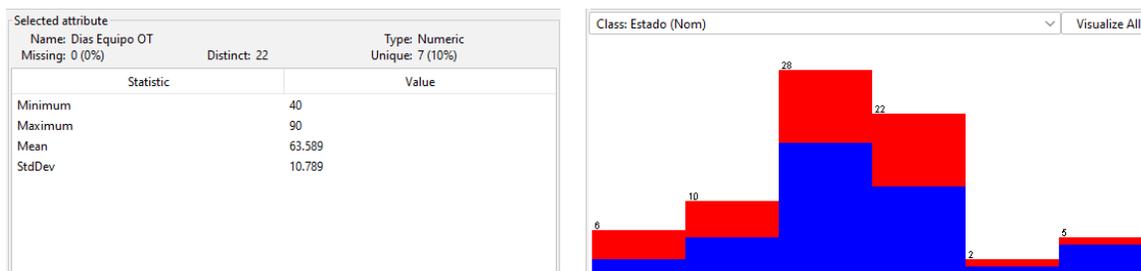
Ejecución de la Orden de Trabajo según la Provincia



La herramienta indica el número de ordenes de trabajo ejecutadas por provincia

Figura 59

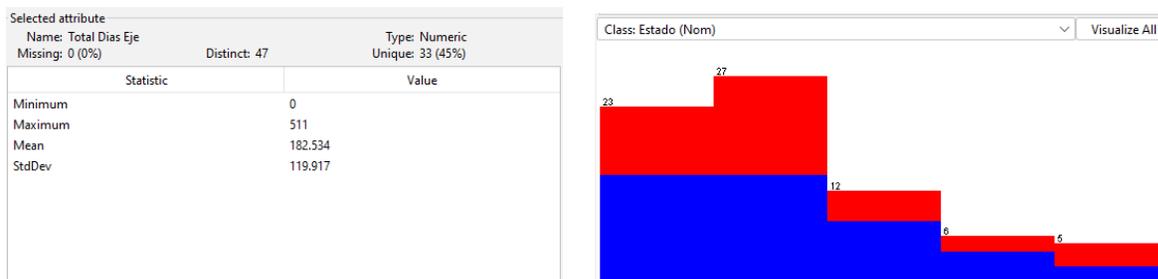
Cantidad de Días de Ejecución de la Auditoría de la Orden de Trabajo



Este indicador representa el número de auditorías de acuerdo con el número de días de la orden de trabajo desde 40 días a 90 días.

Figura 60

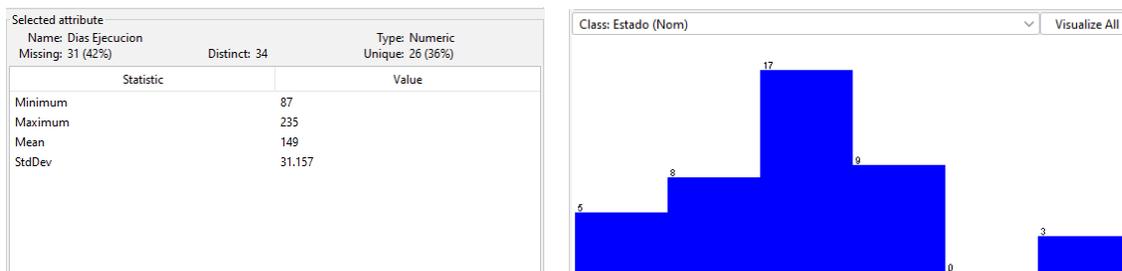
Cantidad de Días de Ejecutados de la Auditoría



Este indicador representa el número de auditorías según el número de días de ejecución desde 0 días a 511 días.

Figura 61

Total de Días de Ejecución del proceso de la Auditoría



Indicador que representa el número de auditorías según el número de días de ejecución desde 87 días a 235 días.

Figura 62

Matriz de Confusión del Modelo de Arboles de Decisión

```

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

Correctly Classified Instances      64      87.6712 %
Incorrectly Classified Instances    9      12.3288 %
Kappa statistic                    0.7766
Mean absolute error                 0.2888
Root mean squared error             0.3412
Relative absolute error             107.8334 %
Root relative squared error         94.1636 %
Total Number of Instances          73

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.923   0.133   0.600     0.923   0.727     0.677   0.968    0.852    Finanzas
                0.750   0.000   1.000     0.750   0.857     0.853   0.981    0.861    Ambiente
                1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    Salud
                0.875   0.040   0.977     0.875   0.923     0.805   0.949    0.954    Educacion
Weighted Avg.   0.877   0.050   0.913     0.877   0.885     0.798   0.959    0.928

=== Confusion Matrix ===
 a  b  c  d  <-- classified as
12  0  0  1 | a = Finanzas
 2  6  0  0 | b = Ambiente
 0  0  4  0 | c = Salud
 6  0  0  42 | d = Educacion

```

Figura 63

Matriz de Confusión del Modelo de Naive Bayes

```

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

Correctly Classified Instances      55          75.3425 %
Incorrectly Classified Instances    18          24.6575 %
Kappa statistic                    0.5297
Mean absolute error                0.1292
Root mean squared error            0.3046
Relative absolute error             48.2515 %
Root relative squared error        84.0648 %
Total Number of Instances          73

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.615   0.117   0.533     0.615   0.571     0.472   0.777    0.551    Finanzas
                0.750   0.062   0.600     0.750   0.667     0.625   0.946    0.815    Ambiente
                0.000   0.014   0.000     0.000   0.000    -0.028   0.377    0.162    Salud
                0.854   0.240   0.872     0.854   0.863     0.609   0.854    0.847    Educacion
Weighted Avg.   0.753   0.186   0.734     0.753   0.742     0.551   0.824    0.754

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
  8  3  0  2 | a = Finanzas
  0  6  0  2 | b = Ambiente
  2  0  0  2 | c = Salud
  5  1  1  4 | d = Educacion

```

Figura 64

Matriz de Confusión del Modelo de JRip

```

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

Correctly Classified Instances      71          97.2603 %
Incorrectly Classified Instances     2          2.7397 %
Kappa statistic                    0.9458
Mean absolute error                0.0149
Root mean squared error            0.1145
Relative absolute error             5.5806 %
Root relative squared error        31.6058 %
Total Number of Instances          73

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    Finanzas
                0.750   0.000   1.000     0.750   0.857     0.853   0.981    0.861    Ambiente
                1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    Salud
                1.000   0.080   0.960     1.000   0.980     0.940   0.992    0.992    Educacion
Weighted Avg.   0.973   0.053   0.974     0.973   0.971     0.944   0.992    0.979

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
 13  0  0  0 | a = Finanzas
  0  6  0  2 | b = Ambiente
  0  0  4  0 | c = Salud
  0  0  0  4 | d = Educacion

```

Figura 65

Matriz de Confusión del Modelo de MultiClassClassifier

```

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

Correctly Classified Instances      73      100 %
Incorrectly Classified Instances    0        0 %
Kappa statistic                    1
Mean absolute error                 0.0001
Root mean squared error             0.0005
Relative absolute error              0.0218 %
Root relative squared error         0.151 %
Total Number of Instances          73

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000   Finanzas
                1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000   Ambiente
                1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000   Salud
                1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000   Educacion
Weighted Avg.   1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
13  0  0  0 | a = Finanzas
 0  8  0  0 | b = Ambiente
 0  0  4  0 | c = Salud
 0  0  0  48 | d = Educacion

```

Tabla 8

Comparación de Resultados según las Matrices de Confusión generadas en los Modelos de Clasificación con Weka con otros campos y con otro banco de datos

Técnica	Correctamente clasificados	Exactitud	Coefficiente Kappa	Clasificados Incorrectamente	Error
Árbol de decisión	64	87,67	0.77	9	12,33
Naive Bayes	55	75.34	0.52	18	24.66
JRip	71	97.26	0.94	2	2.74
MultiClass Classifier	73	100	1	0	0

Figura 66

Modelo: Naive Bayes Classifier - Categorizado por Unidad

```

=== Summary ===
Correctly Classified Instances      47          64.3836 %
Incorrectly Classified Instances    26          35.6164 %
Kappa statistic                    0.5554
Mean absolute error                0.0683
Root mean squared error            0.2296
Relative absolute error            44.5501 %
Root relative squared error        83.1837 %
Total Number of Instances          73

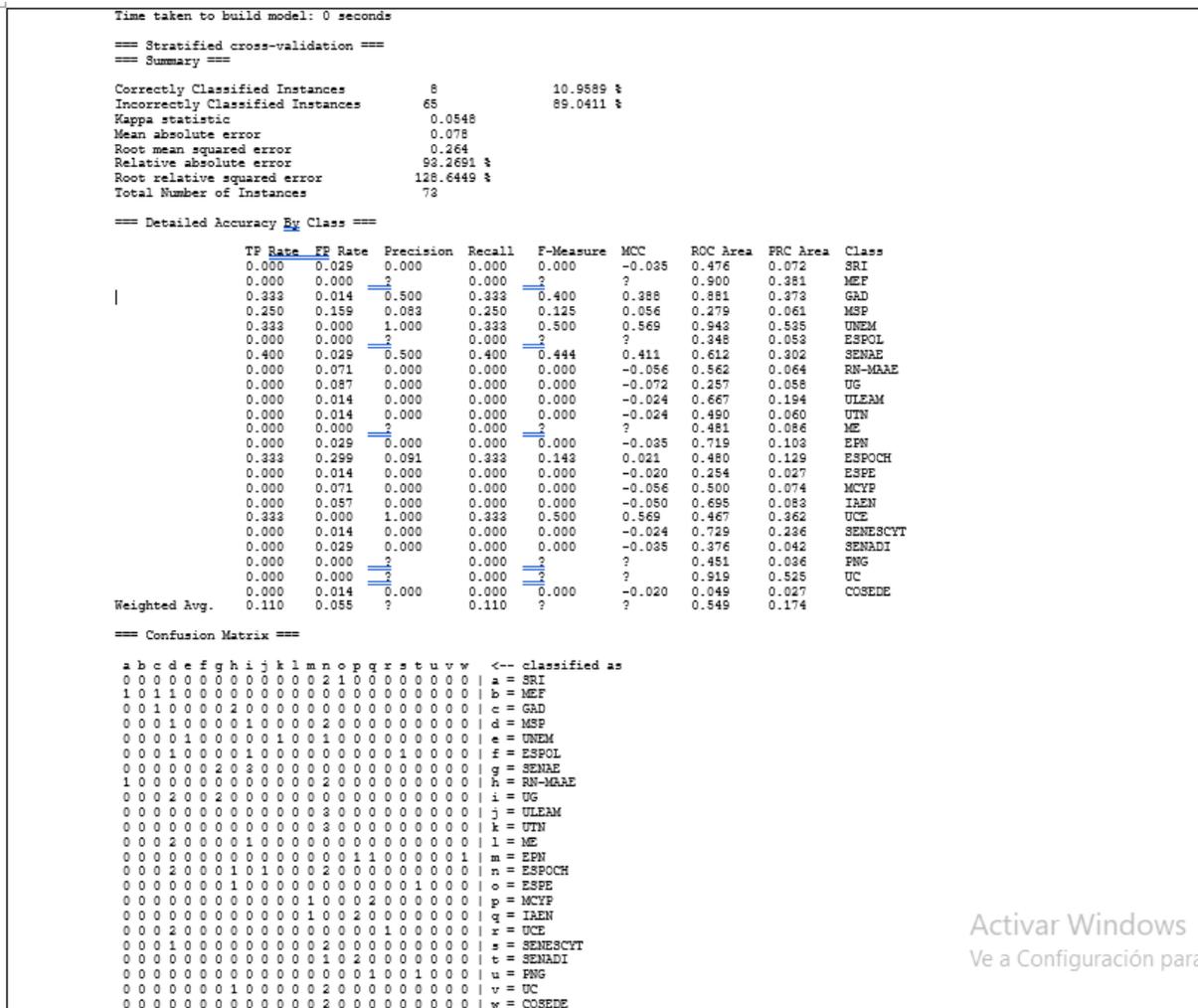
=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.625  0.062  0.556  0.625  0.588  0.595  0.844  0.691  DNA3
0.333  0.000  1.000  0.333  0.500  0.569  0.957  0.750  DNA5
0.250  0.014  0.500  0.250  0.333  0.328  0.409  0.164  DNA7
1.000  0.034  0.882  1.000  0.938  0.923  0.983  0.925  DPGY
0.000  0.086  0.000  0.000  0.000  -0.062  0.543  0.064  DNA6
0.000  0.000  ?      0.000  ?      ?      0.676  0.459  DPM-
0.000  0.000  ?      0.000  ?      ?      0.529  0.065  DPI-
1.000  0.100  0.821  1.000  0.902  0.860  0.996  0.992  DNA2
0.333  0.119  0.200  0.333  0.250  0.171  0.614  0.439  DPCH
0.000  0.000  ?      0.000  ?      ?      0.458  0.037  DPG-
0.000  0.000  ?      0.000  ?      ?      0.900  0.545  DFA-
Weighted Avg.  0.644  0.059  ?      0.644  ?      ?      0.842  0.702

=== Confusion Matrix ===
 a  b  c  d  e  f  g  h  i  j  k  <-- classified as
5  0  0  0  2  0  0  0  1  0  0  | a = DNA3
0  1  0  0  2  0  0  0  0  0  0  | b = DNA5
1  0  1  2  0  0  0  0  0  0  0  | c = DNA7
0  0  0  15  0  0  0  0  0  0  0  | d = DPGY
1  0  0  0  0  0  0  0  2  0  0  | e = DNA6
0  0  0  0  0  0  0  0  3  0  0  | f = DPM-
0  0  0  0  0  0  0  0  3  0  0  | g = DPI-
0  0  0  0  0  0  0  23  0  0  0  | h = DNA2
2  0  1  0  1  0  0  0  2  0  0  | i = DPCH
0  0  0  0  0  0  0  1  1  0  0  | j = DPG-
0  0  0  0  0  1  0  0  2  0  0  | k = DFA-

```

Figura 67

Modelo: Naive Bayes Classifier - Categorizado por Entidad



Activar Windows
Ve a Configuración para

Figura 68

Modelo: Naive Bayes Classifier - Categorizado por Entidad

```

==== Classifier model (full training set) ====
Naive Bayes Classifier

Attribute      Class
              Financas  Ambiente  Salud  Educacion
              (0.15)    (0.12)    (0.06)  (0.64)
-----
Cod OT
mean          26455.8846  26573.625  26415.625  26561.2813
std. dev.     1076.319   987.5698   1134.68   1076.8005
weight sum    13         8          4         48
XXXXXXXXXX    39.5      39.5      39.5      39.5

CodIME
DIA3          9.0        1.0        1.0        1.0
DIA5          1.0        4.0        1.0        1.0
DIA7          1.0        1.0        5.0        1.0
DPFY         6.0        1.0        1.0        11.0
DIA6          1.0        4.0        1.0        1.0
DFM-          1.0        1.0        1.0        4.0
DFI-          1.0        1.0        1.0        4.0
DIA2          1.0        1.0        1.0        24.0
DFCH         1.0        1.0        1.0        7.0
DFG          1.0        3.0        1.0        1.0
DFA-         1.0        1.0        1.0        4.0
[total]      24.0      19.0      15.0      59.0

CodEnte
SRI           4.0        1.0        1.0        1.0
MEF           4.0        1.0        1.0        1.0
GAD           1.0        4.0        1.0        1.0
MSP           1.0        1.0        5.0        1.0
UNEM          1.0        1.0        1.0        4.0
ESPOL         1.0        1.0        1.0        4.0
SENAE         6.0        1.0        1.0        1.0
RM-MAAE      1.0        4.0        1.0        1.0
UG           1.0        1.0        1.0        5.0
ULEAM         1.0        1.0        1.0        4.0
UTN           1.0        1.0        1.0        4.0
ME            1.0        1.0        1.0        4.0
EPN           1.0        1.0        1.0        4.0
ESFOCH        1.0        1.0        1.0        7.0
ESPE         1.0        1.0        1.0        3.0
MCYP         1.0        1.0        1.0        4.0
IAEN         1.0        1.0        1.0        4.0
UCE           1.0        1.0        1.0        4.0
SENESCYT     1.0        1.0        1.0        4.0
SENADI       1.0        1.0        1.0        4.0
PHG          1.0        3.0        1.0        1.0
UC           1.0        1.0        1.0        4.0
COSEDE       3.0        1.0        1.0        1.0
[total]      36.0      31.0      27.0      71.0

Provincia
EJCINCHA     9.0        7.0        5.0        24.0
GUAYAS       6.0        1.0        1.0        11.0
MANABI       1.0        1.0        1.0        4.0
IMBABURA     1.0        1.0        1.0        4.0
CHIMBORAZO  1.0        1.0        1.0        7.0
GALAPAGOS   1.0        3.0        1.0        1.0
AZUAY        1.0        1.0        1.0        4.0
[total]      20.0      15.0      11.0      55.0

```

```

Dias Equipo OT
mean      69.2308  67.8571  65.0476  60.4167
std. dev. 12.7181  1.6936  17.4152  9.0488
weight sum      13      8      4      48
precision      2.381  2.381  2.381  2.381

Total Dias Eje
mean      243.5368  133.3043  249.9457  167.3247
std. dev. 151.4687  38.4817  31.9073  113.636
weight sum      13      8      4      48
precision      11.1087  11.1087  11.1087  11.1087

Dias Equipo H
mean      1344.8637  1829.4814  1773.5903  1592.0417
std. dev. 655.6585  124.1261  288.5782  495.6928
weight sum      13      8      4      48
precision      55.8611  55.8611  55.8611  55.8611

Dias Equipo Ses
mean      20.5882  10.2941  20.5882  10.7059
std. dev. 27.851  2.9116  11.6465  18.2908
weight sum      6      3      3      30
precision      6.1765  6.1765  6.1765  6.1765

Dias Equipo Region
mean      159.9596  201.8182  131.5556  143.2162
std. dev. 33.2269  29.9736  26.4906  24.2329
weight sum      6      3      3      30
precision      4.4848  4.4848  4.4848  4.4848

Estado
Completo      7.0      4.0      4.0      31.0
Incompleto    8.0      6.0      2.0      19.0
Total         15.0     10.0     6.0      50.0

```

Time taken to build model: 0 seconds

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

```

Correctly Classified Instances      55      75.3425 %
Incorrectly Classified Instances    18      24.6575 %
Kappa statistic                    0.5297
Mean absolute error                 0.1292
Root mean squared error             0.3046
Relative absolute error             48.2515 %
Root relative squared error        84.0648 %
Total Number of Instances          73

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.615	0.117	0.533	0.615	0.571	0.472	0.777	0.551	Finanzas
	0.750	0.062	0.600	0.750	0.667	0.625	0.946	0.815	Ambiente
	0.000	0.014	0.000	0.000	0.000	-0.028	0.377	0.162	Salud
	0.854	0.240	0.872	0.854	0.863	0.609	0.854	0.847	Education
Weighted Avg.	0.753	0.186	0.734	0.753	0.742	0.551	0.824	0.754	

=== Confusion Matrix ===

```

a b c d <-- classified as
2 0 0 2 | a = Finanzas
0 6 0 2 | b = Ambiente
2 0 0 2 | c = Salud
5 1 1 4 | d = Education

```

Figura 69

Modelo: MultiClass Classifier - Categorizado por Ámbito

=== Stratified cross-validation ===

=== Summary ===

```

Correctly Classified Instances      73      100 %
Incorrectly Classified Instances    0      0 %
Kappa statistic                    1
Mean absolute error                 0.0001
Root mean squared error             0.0005
Relative absolute error             0.0218 %
Root relative squared error        0.151 %
Total Number of Instances          73

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Finanzas
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Ambiente
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Salud
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Education
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

```

a b c d <-- classified as
13 0 0 0 | a = Finanzas
0 8 0 0 | b = Ambiente
0 0 4 0 | c = Salud
0 0 0 48 | d = Education

```

Capítulo IV

Análisis de Resultados

Interpretación y Evaluación

Para el presente estudio se propuso los siguientes modelos de clasificación: árboles de decisión, Naive Bayes, JRip y MultiClass Classifier, ZeroR. La comparación de los cuatro modelos nos permite evaluar cada uno de ellos y poder determinar cuál de los modelos planteados, el que mejor se adapta a nuestro caso de estudio, el que más aciertos de predicción tenga, así como el menor porcentaje de error, la correlación que existe entre el tipo de examen ejecutado y el tiempo de cumplimiento.

El algoritmo de clasificación de Naïve Bayes, es un algoritmo de aprendizaje automático supervisado, que se lo utiliza para las tareas de clasificación, como la clasificación de texto. También es parte de una familia de algoritmos de aprendizaje generativo, lo que significa que busca modelar la distribución de entradas de una clase o categoría determinada

El modelo ZeroR es el algoritmo de clasificación que se basa en el objetivo e ignora todos los predictores. El clasificador ZeroR es el mas sencillo, pronostica la categoría mayoritaria de la clase, este modelo no tiene poder de previsibilidad, pero es útil para determinar un rendimiento como punto de referencia para otros modelos de clasificación.

La clasificación multiclase empleado en el aprendizaje automático y la clasificación estadística, se lo emplea para solucionar problemas que requieren clasificar instancias de tres o más clases.

Así mismo se realizó la evaluación de resultados con otros campos y otro banco de datos para los modelos de predicción y clasificación arboles de decisión y Naive Bayes, mediante el análisis e interpretación de ellos. Los dos modelos leen el mismo archivo de datos, toman la información en una relación del 80% para el aprendizaje del modelo y el 20% restante para las pruebas, a través de la variable Examen Cumplido se determinó la predicción del

modelo, determinando que el modelo de Naive Bayes para el presente estudio indica mejores resultados.

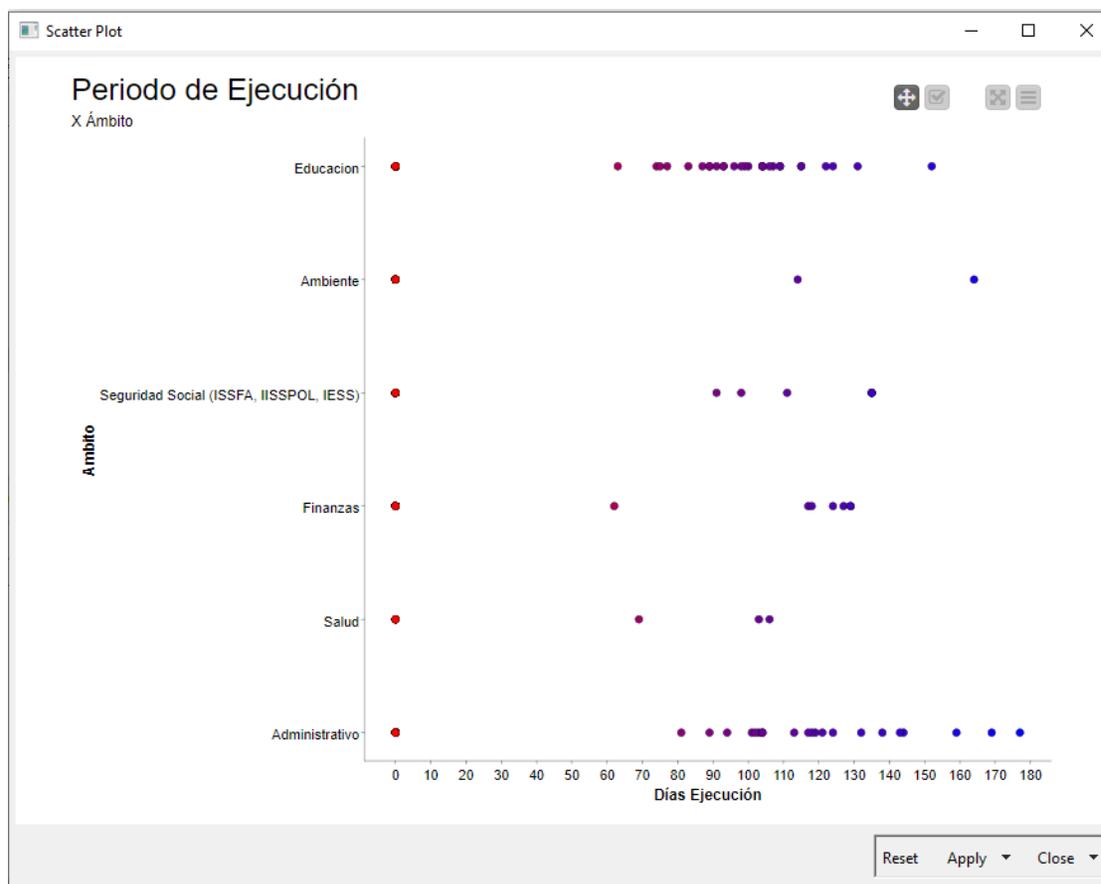
La comparación de los dos modelos de clasificación con otro banco de datos y otros campos permite evaluar cada uno de ellos y poder determinar cuál de los modelos planteados es el que mejor se adapta a nuestro caso de estudio, el que más aciertos de predicción tenga, así como el menor porcentaje de error, se observa que cuando existe una buena correlación de los datos, por ejemplo, la correlación entre el tipo de examen ejecutado y el tiempo de cumplimiento, el modelo Naive Bayes alcanza una mayor exactitud.

Así mismo se observa, el coeficiente kappa para el modelo de naive bayes es 1, lo que significa que hay un buen grado de concordancia inter-observador. Para el modelo de árbol de decisión, el valor de $k = 0$ refleja que la disconcordancia observada es precisamente lo que se espera a causa exclusivamente del azar.

En la Figura 70 se muestran los resultados de ejecución del modelo K-means

Figura 70

Resultado del Gráfico de Dispersión (Nodo Scatter Plot) del Modelo K-means



En base a las figuras anteriores, se realiza una comparación entre las matrices de confusión, tal y como se muestra en la siguiente Tabla 13.

De los resultados observados a la ejecución de los 2 modelos de árbol de decisión se observa que el modelo de naive bayes es el que mejor se adapta para el caso de estudio.

El modelo de K-Means se incorporó el nodo Scatter Plot, que permite evaluar los resultados a través de un gráfico de dispersión, en el que se puede observar que la mayor concentración de los puntos de un plano cartesiano, se ubican en el rango de los [80, 140] días.

Se puede indicar que existe una correlación fuerte en el rango de [80, 140] días, una correlación débil en los rangos de [60, 79] días y [141, 180] días, y ninguna correlación en el rango de [0, 59] y para un tiempo mayor a 180 días.

Implantación y Despliegue

En esta fase se explota el conocimiento adquirido a través del modelo de árbol de decisión generado y k-means, a continuación, se describen las actividades de esta fase.

Para realizar el despliegue del modelo generado es necesario seguir los siguientes pasos:

Permisos de consulta a las tablas a los esquemas de PlanControl y Control

Ejecución del script de consulta en la base de datos Sql Server "scriptModelos.txt"

Los resultados de la ejecución copiar y trasladar a una hoja de cálculo en formato Excel y guardarlo.

Descargar de la web la herramienta Knime Analytics Platform vs 4.6.1 o superior e instalarla.

En el modelo de árbol de decisión, como en el modelo k-means a ejecutar, configurar en el primer nodo, "Excel Reader" para que lea el archivo generado de la consulta "scriptModelo.txt"

Ejecutar todos los nodos del modelo de árbol de decisión de la herramienta knime generado y verificar el comportamiento del modelo.

Ejecutar todos los nodos del modelo de K-means de la herramienta knime generado y verificar el comportamiento del modelo.

Para el mantenimiento de los modelos de árboles de decisión y k-means generados, se debe considerar la frecuencia y periodicidad con la que se actualiza la información del archivo generado "scriptModelos.txt", script de consulta, tomando en cuenta que la información de los sistemas transaccionales es en línea, sin embargo, dependen del registro de los datos en dichos sistemas. Los dueños del proceso de seguimiento deben definir la periodicidad de la ejecución del script de consulta que permite generar y actualizar los datos en el archivo de inicio de ejecución de los modelos, de acuerdo a las necesidades que la Entidad de Control lo requiera.

En los resultados de la ejecución del modelo K-means, se observa el segmento que arroja dicho modelo para la ejecución de los exámenes de auditoría, aquí se observan exámenes que tiene una ejecución temprana y otros que tienen una ejecución más demorada que casi alcanzan el tope del tiempo según la normativa vigente; y es a éstas ejecuciones de auditoría que se debe realizar gestión en coordinación con la unidad ejecutora ya sea con ampliación del tiempo o con la asignación de personal de apoyo si fuese el caso.

La Unidad administrativa que tiene a su cargo la Administración Funcional de la función de control y auditoría debe analizar el periodo de tiempo de actualización del archivo, para de esta manera tener una visión temprana del comportamiento de los equipos ejecutores de las ordenes de trabajo y así realizar un seguimiento continuo para ayudar a resolver novedades que se presenten al equipo ejecutor y observar el registro de los datos en los sistemas transaccionales, considerar un periodo de tiempo adecuado ya que puede verse reflejado en el tiempo de ejecución del examen de auditoría.

Capítulo V

Conclusiones

Se ha logrado estudiar y proponer un modelo analítico-predictivo que aplicando varias herramientas de análisis de datos y técnicas de minería de datos se han encontrado diferentes resultados en los niveles de cumplimiento mediante la categorización de los indicadores por tipo de examen y unidad que tiene establecido la entidad de control.

Se realizó la revisión de literatura para determinar las técnicas predictivas más utilizadas de Machine Learning y apropiadas para el organismo de control. Esto permitió seleccionar los indicadores sensibles en la operación auditada de la entidad.

Se ha verificado la calidad y consistencia de los datos previo a ser utilizados en el modelo predictivo.

Se aplicaron las técnicas de minería de datos y una metodología de desarrollo con el fin de ejecutar el proceso completo, desde la selección de la data, análisis, depuración, selección, estudio y definición de patrones, previo a la aplicación de diferentes modelos de predicción y herramientas de apoyo.

Se han realizado diferentes pruebas, análisis de la data y sus resultados hasta lograr la implementación de un modelo analítico-predictivo que permita construir modelos basados en los factores e indicadores identificados que en forma posterior permita tener una visión a corto plazo de la operación de la entidad auditada, su estado de cumplimiento, y establecer controles a los equipos de trabajo, el seguimiento y cumplimiento de planes de trabajo, entre otros que en este documento no ha sido posible publicar debido a la confidencialidad de la data principal.

Se ha realizado un estudio completo y se han obtenido resultados que contribuyen a la entidad en la implementación de controles, sin embargo, en este documento no se encuentran publicados todos los resultados, debido a la naturaleza de la entidad y la confidencialidad de los datos, no se debe publicar.

También, se ha validado el modelo seleccionado, a través los resultados obtenidos, de patrones y tendencias de los indicadores, y al compararlos con resultados obtenidos de forma manual, se ha determinado un nivel de confianza significativo.

Recomendaciones

De la revisión y validación con los funcionales de las dos herramientas, se observó que la planificación una vez aprobada, es ejecutada en el siguiente proceso de control, para iniciar con el estudio y ejecución de la orden de trabajo, sin embargo, por razones fuera del alcance institucional se dan planificaciones posteriores o se generan exámenes imprevistos, por lo que se recomienda que dichos registros sean subidos en el proceso de control de manera individual e integrada.

En el registro de los datos de la planificación, el sistema debe validar los campos obligatorios y no permitir el guardado con campos vacíos, debe presentar la lista de posibles valores para que el usuario seleccione datos válidos.

Los modelos construidos en el presente estudio son el inicio para realizar estudios cada vez más profundos y se tengan resultados de acuerdo a las necesidades de la institución.

Bibliografía

Apolloni, B., Ghosh, A., Alpaslan, F., & Patnaik, S. (2005). *Machine learning and robot perception, Vol 7*. Madrid: Springer Science & Business Media.

Asamblea Nacional del Ecuador. (20 de Octubre de 2008). CONSTITUCIÓN DE LA REPUBLICA DEL ECUADOR. *CONSTITUCIÓN DE LA REPUBLICA DEL ECUADOR*, 44. Ecuador.

Asamblea Nacional del Ecuador. (2008). Constitución de la República del Ecuador. *Constitución de la República del Ecuador, art. 211 y 212, 44*.

- Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL – RTE, Vol. 29, N. 1, 170-183*. Obtenido de <http://200.10.150.204/index.php/tecnologica/article/view/464>
- CGE, Acuerdo 016-CG-2020. (11 de 08 de 2020). REGLAMENTO DE ÁMBITO DE CONTROL DE LAS UNIDADES ADMINISTRATIVAS DE LA CONTRALORÍA GENERAL DEL ESTADO. *Acuerdo 016-CG-2020*. Obtenido de <https://www.contraloria.gob.ec/Normatividad/BaseLegal>
- CGE, Acuerdo 056-CG-2018. (2018). ESTATUTO ORGANICO, Acuerdo 056-CG-2018.
- CGE, Acuerdo 057 CG 2018. (2018). Estatuto Orgánico Sustitutivo Acuerdo 057 CG 2018. Quito: CGE.
- Cleophas, T., Zwinderman, A., & Cleophas, H. (2013). *Machine learning in medicine*. Springer, Dordrecht. Obtenido de https://doi.org/10.1007/978-94-007-6886-4_1
- Contreras Barrera, M. (2016). Minería de texto en la clasificación de material bibliográfico. *redalyc.org*. Obtenido de <https://www.redalyc.org/journal/161/16148511003/html/>
- DataScientest. (03 de 2022). *datascientest.com*. Obtenido de <https://datascientest.com/es/text-mining-o-mineria-de-textos-definicion-tecnicas-casos-de-uso>
- Datos, E. (2021). <https://elmundodelosdatos.com/introduccion-al-clustering-algoritmo-k-means/>.
- Dávila, E. (2018). La inteligencia artificial fue un potente aliado en el 2018. Obtenido de <https://www.elcomercio.com/guaifai/inteligencia-artificial-potente-aliado-2018.html>
- El Naqa&Murphy. (2015). *What is machine learning? Machine learning in radiaton*. Obtenido de https://doi.org/10.1007/978-3-319-18305-3_1
- Estatuto Orgánico de Gestión Organizacional por Procesos. (30 de Julio de 2020). *Acuerdo 015-CG-2020*.
- Fayyad, U., Gregory Piatetsky-Shapiro, & Padhraic Smyth. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*.

- Flores, A. (06 de 2021). *crehana*. Obtenido de <https://www.crehana.com/blog/transformacion-digital/herramientas-mineria-datos/>
- Garcia, O. A. (2021). *Aplicación del sistema de Machine Learning para aumentar la eficiencia de las organizaciones*. Obtenido de <http://hdl.handle.net/10654/39730>
- Gong, Y., & Xu, W. (2007). *Machine learning for multimedia content analysis (Vol. 30)*. . New York: Springer Science & business media.
- Györfi, L., Ottucsák, G., & Walk, H. (2012). *Machine learnig for Financial engineering (Vol. 8)*. London: World Scientific.
- Herrera, M., Ruiz, S., Romagnano, M., & Lund, M. I. (2017). Determinación de Variables que Influyen en el Rendimiento Académico. *Libro de Actas: XX Encuentro Nacional y XII Internacional Educación Matemática en Carreras de Ingeniera*, 337-346. Obtenido de https://www.researchgate.net/profile/Maria-Ines-Lund/publication/358151255_Determinacion_de_Variables_que_Influyen_en_el_Rendimiento_Academico/links/61f339e18d338833e39ceceb/Determinacion-de-VARIABLES-que-Influyen-en-el-Rendimiento-Academico.pdf
- IONOS. (05 de 2022). *IONOS*. Obtenido de <https://www.ionos.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas/>
- Jan, C. L. (2021). Using Deep Learning Algorithms for CPAs' Going Concern Prediction. *Information*.
- Koskivaara, E. (2017). Artificial neural network models for predicting patterns in auditing monthly balances. *Journal of the Operational Research Society*.
- Landa, J. (2016). *¿Qué es KDD y Minería de Datos?* Obtenido de <https://fcojlanda.me/es/ciencia-de-los-datos/kdd-y-mineria-de-datos-espanol/>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444. Obtenido de <https://www.nature.com/articles/nature14539>

- León Guzmán, E. (2016). *Universidad Nacional de Colombia*. Obtenido de https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf
- Ley Orgánica CGE. (12 de Junio de 2002). *Ley Orgánica, arts. 18, 19, 24, art. 26*.
- Melillanca, E. (08 de 2018). *Evaluación de modelos de clasificación*. Obtenido de <http://www.ericmelillanca.cl/content/evaluaci-n-modelos-clasificaci-n-matriz-confusi-n-y-curva-roc>
- Microsoft. (2022). *Analysis Services - Minería de datos*. Obtenido de Analysis Services - Minería de datos: <https://docs.microsoft.com/es-es/analysis-services/data-mining/cross-validation-analysis-services-data-mining?view=asallproducts-allversions>
- Microsoft. (2022). *Matrices de confusión*. Obtenido de Matrices de confusión: <https://docs.microsoft.com/es-es/learn/modules/machine-learning-confusion-matrix/2-confusion-matrices>
- Mohammadi, M., Yazdani, S., Khanmohammadi, M., & Maham, K. (2020). Financial reporting fraud detection: An analysis of data mining algorithms. *International Journal of Finance & Managerial Accounting*.
- Molina López, J. M., & García Herrero, J. (2006). Obtenido de TÉCNICAS DE ANÁLISIS DE DATOS: http://matema.ujaen.es/jnavas/web_recurso/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- Montenegro Fierro, W. J. (2020). Modelo de fidelización para reducir la cancelación de los servicios que ofrece una empresa de telecomunicaciones.
- Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). Las implicaciones éticas del uso de la inteligencia artificial en la auditoría. *J Bus Ética 167, 209–234 (2020)*., <https://doi.org/10.1007/s10551-019-04407-1>.
- Muslihatun, F. A., Hantono, B. S., & Fauziati, S. (2020). Continuous auditing and data mining for strategic risk control and anticorruption: Creating “fair” value in the digital age.

- Nesreen, Amir, Neamat, & Hisham. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting .
- Oded, M., & Lior, R. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Ordoñez, H., Cobos, C., & Bucheli, V. (2020). Machine learning model for predicting theft trends in Colombia. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*.
- Rahayu, S. (2021). Utilization of artificial intelligence in tax audit in indonesia. Obtenido de <https://mar.uitm.edu.my/images/Vol-20-3/06.pdf>
- Rahman, R., Masrom, S., Zakaria, N., & Halid, S. (2021). Auditor Choice Prediction Model Using Corporate Governance and Ownership Attributes: Machine Learning Approach.
- Ramos, X. (19 de Agosto de 2018). La era del robot se instala a paso lento en Ecuador. pág. 3.
- Raschke, R. L., Saiewitz, A., Kachroo, P., & Lennard, J. B. (2018). Consulta de auditoría mejorada con IA: una nota de investigación. <https://doi.org/10.2308/jeta-52310>.
- Reyes, A. (02 de 2020). *Machine Learning: Qué es, para qué sirve*. Obtenido de Machine Learning: <https://aticaingenieria.cl/machine-learning-que-es-para-que-sirve/>
- Rueda, J. F. (2022). *Data mining*. Obtenido de Data mining: <https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/>
- Sancho Caparrini, F. (12 de 2020). *Aprendizaje Supervisado y No Supervisado*. Obtenido de <http://www.cs.us.es/~fsancho/?e=77>
- Schwarz, M. (01 de 2013). *Gestión de Operaciones y Proyectos Mineros*. Obtenido de <http://max-schwarz.blogspot.com/2013/01/marco-teorico-vs-estado-del-arte-en-la.htm/>
- Sequeira Ortíz, Z. (2019). Tendencias y desafíos para las ciencias de la información en el mundo actual. *E-Ciencias de la Información*, 9(1), 196-208. Obtenido de https://www.scielo.sa.cr/scielo.php?pid=S1659-41422019000100196&script=sci_arttext
- She-I Chang, C.-F. T.-H., & Hwang, C. L. (2008). The development of audit detection risk assessment system: Using the fuzzy theory and audit risk model.

- Silva, D. d. (12 de 08 de 2021). *Qué es el Deep Learning?* Obtenido de Web Content & SEO Associate, LATAM: <https://www.zendesk.com.mx/blog/que-es-el-deep-learning/>
- Sun, T. (2019). Aplicación del aprendizaje profundo a los procedimientos de auditoría: un marco ilustrativo. . *Horizontes contables* , 33 (3), 89-109., <https://doi.org/10.2308/acch-52455>.
- Sun, T., & Vasarhelyi, M. (2017). Auditor Choice Prediction Model Using Corporate Governance. *CPA Journal*.
- Sun, T., & Vasarhelyi, M. (2017). Deep Learning and the Future of Auditing: How an Evolving Technology Could Transform Analysis and Improve Judgment. *CPA Journal*.
- Timarán-Pereira, S. R.-A.-Z.-T. (2016). *El proceso de descubrimiento de conocimiento en bases de datos*. Obtenido de <http://dx.doi.org/10.16925/9789587600490>
- Toapanta, S. M., Ortiz, S. E., & Gallegos, L. E. (2020). Definición de un Prototipo para la Realización de Auditorías Informáticas en una Organización Pública del Ecuador. Universidad Nacional de La Matanza. (2020). Proceso de Design Science Research. *Revista Digital del Departamento de Ingeniería e Investigaciones Tecnológicas*.
- UTI. (2019). Universidad Tecnológica Indoamérica realizó conferencia magistral de inteligencia artificial. *Conferencia Magistral: Inteligencia Artificial más allá de los límites legales*. Obtenido de [uti.edu.ec::http://www.uti.edu.ec/~utiweb/universidad-indoamerica-realizo-conferenciamagistral-inteligencia-artificial-mas-alla-de-los-limites-legales/](http://www.uti.edu.ec/~utiweb/universidad-indoamerica-realizo-conferenciamagistral-inteligencia-artificial-mas-alla-de-los-limites-legales/)