# *Modelo de pronósticos para indicadores operativos de auditoria en Entidad de Control, basado en las técnicas de Machine Learning.*

Autora: Carmen Isabel García Llanos

Directora: Ing. Sonia Cárdenas D., Ph.D.

# Agenda

- Antecedentes
- Objetivos
- Metodología
- Desarrollo
- Análisis de resultados
- Conclusiones
- Recomendaciones

# Antecedentes

**Problema**

No cuenta con un modelo que aplique técnicas de aprendizaje automático

**Justificación**

Evaluar y medir el desempeño de los procesos relacionados con la función de auditoría.

**Importancia**

Considerar el estudio como un prototipo de análisis predictivo para las necesidad institucionales.

**Alcance**

Encontrar el mejor modelo para contribuir en la toma de decisiones.

# 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.

# Objetivos Específicos

- Realizar una revisión de la literatura para determinar las técnicas predictivas más utilizadas de Machine Learning y apropiadas para el organismo de control.

- Seleccionar los indicadores sensibles en la operación auditada de la entidad.

- Verificar la calidad y consistencia de los datos que van a ser utilizados en el modelo predictivo, aplicando la metodología de desarrollo de procesos de minería de datos.

# Objetivos Específicos

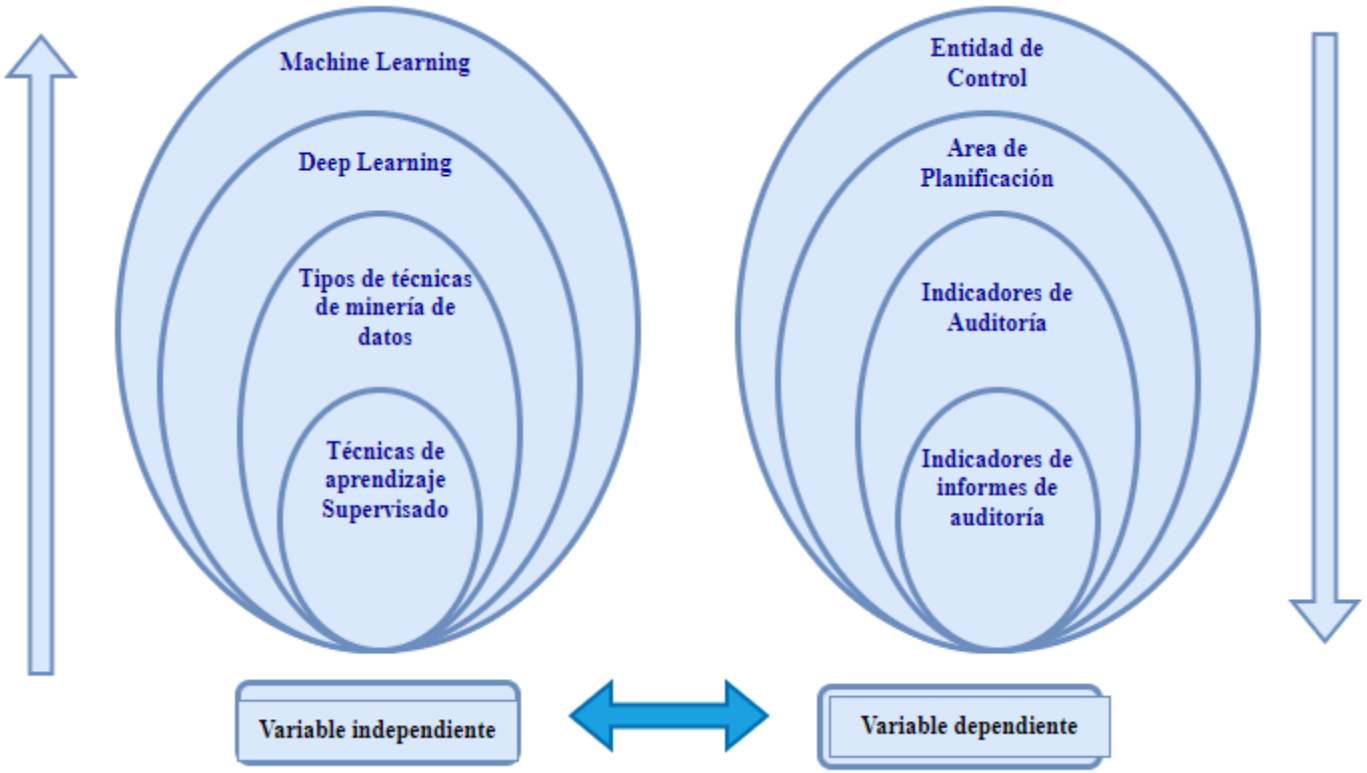
- Implementar un modelo analítico-predictivo mediante técnicas de ML para construir modelos predictivos basados en indicadores de auditoría, que permita tener una visión a corto plazo de la operación de la entidad auditada.

- Validar el modelo predictivo, a través 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.

# Hipótesis

La aplicación de las técnicas de Machine Learning permitirán tener las predicciones en los indicadores de auditoría de la Entidad de Control.

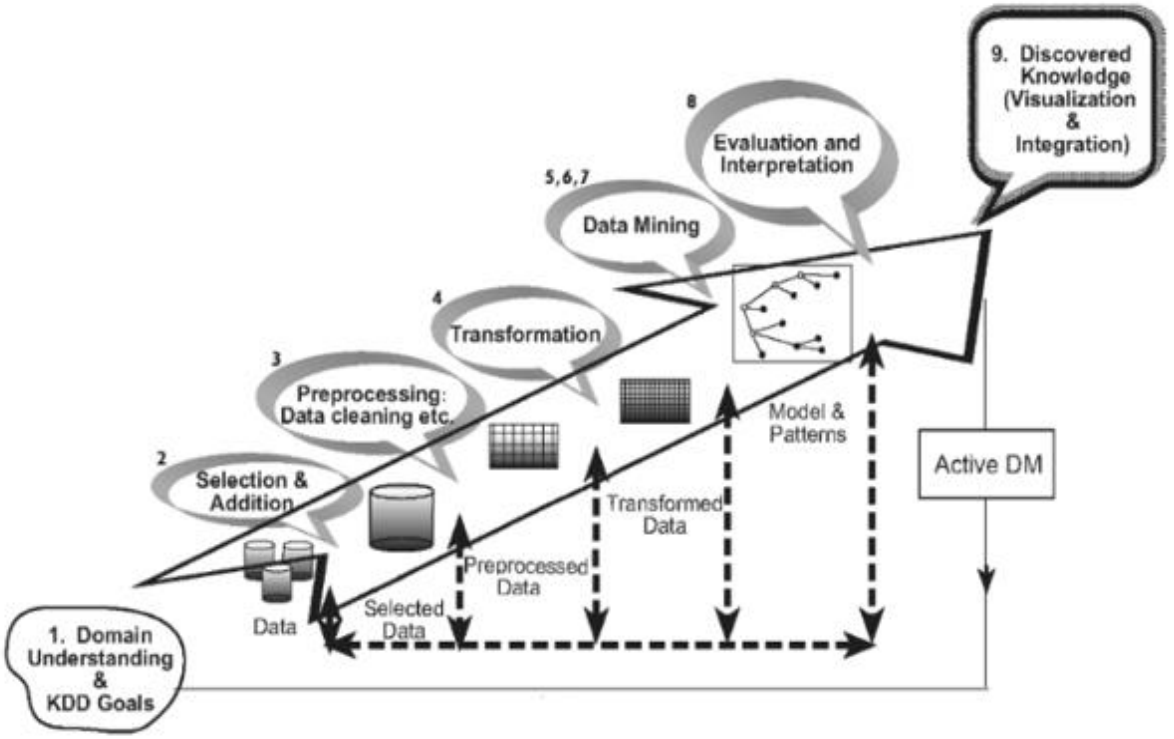
# Marco teórico



# Metodología

## Knowledge Discovery in Databases (KDD)

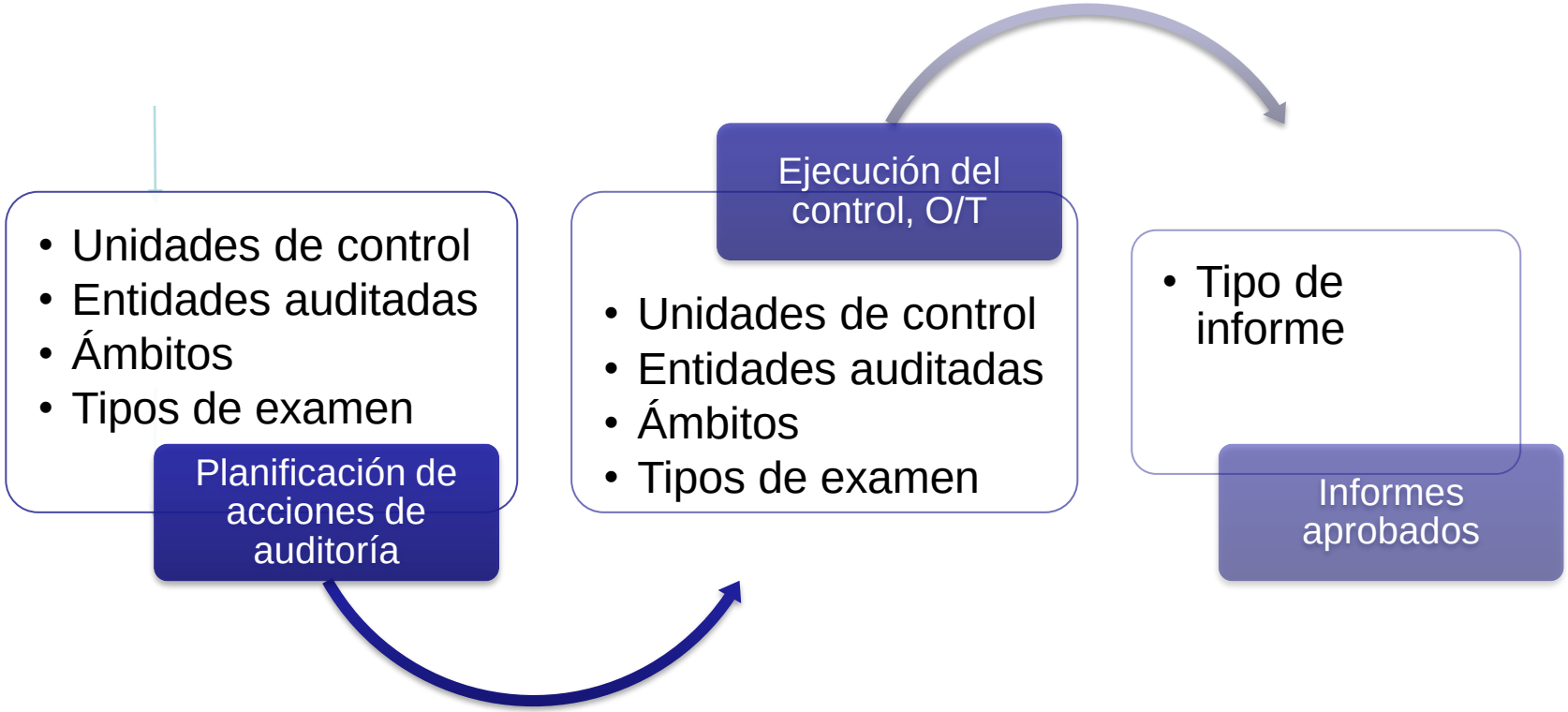
- Selección,
- Preprocesamiento,
- Transformación,
- Minería de datos y
- Evaluación e interpretación



# Selección

# Desarrollo

## Procesos seleccionados para el Estudio



# Desarrollo

## Selección de datos

*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 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])
```

133 %

Results | Messages

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

# Desarrollo

## Selección de datos

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)
```

133 %

Results | Messages

| | 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-2022 |
| 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-2022 |
| 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-TVIPyA-DGA... |

# Desarrollo

## Selección de datos

*Datos **seleccionados de la** Ejecución **del Control del año 2022***

| ordtraId | ordtraFechaRegistro | NumeroOrden Trabajo | Objetivos | Alcance | EntidadAudi | UnidadAudit | EntidadExam | LugarAccion | FechaOrden | detperIdFirn | tipexaId | PeriodoExan | PeriodoExan |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25086 | 2022-01-03 11:10:52.1 | 0001-DPI-AE-2022 | Los objetivo: | a los gastos; | 12963 | 9291 | 32665 | ATUNTAQUI | 3/1/2022 | 49586 | 13 | 2/1/2017 | 31/12/2021 |
| 25087 | 2022-01-03 11:15:43.2 | 0002-DPI-AE-2022 | Los objetivo: | 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 objetivo: | 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 objetivo: | 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 objetivo: | 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 11:21:14.3 | 0005-DPI-AE-2022 | Los objetivo: | 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 objetivo: | 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 objetivo: | 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 11:29:42.0 | 0001-DNA5-GAD-EPMHV-A | Los objetivo: | a los ingreso | 29429 | 0 | 29429 | Empresa Púb | 3/1/2022 | 47833 | 13 | 2/1/2018 | 31/12/2021 |
| 25096 | 2022-01-03 11:30:19.2 | 0001-DNA8-TVIPyA-DGAC-A | Los objetivo: | a las fases pr | 11638 | 0 | 11638 | Quito | 3/1/2022 | 50393 | 13 | 1/5/2019 | 31/7/2021 |
| 25097 | 2022-01-03 11:33:54.2 | 0008-DPI-AE-2022 | Los objetivo: | a los ingreso | 12963 | 9291 | 34102 | IBARRA | 3/1/2022 | 4/10/2035 | 13 | 2/1/2017 | 31/12/2021 |
| 25098 | 2022-01-03 11:42:06.9 | 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 11:48:36.9 | 0001-DNA2-MIDUVI-AI-202 | Los objetivo: | 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 objetivo: | al cumplimie | 27426 | 0 | 27426 | Ibarra | 3/1/2022 | 16/3/2036 | 13 | 1/6/2017 | 31/12/2021 |

# Desarrollo

## Selección de datos

***Consulta de los Informes Aprobados del Año 2022***

```
--tiempo de ejecución de la orden de trabajo y la aprobación del informe
-- encontrar los mejores tiempos de ejecución de la orden de trabajo
--segun los ambitos
select i.infNumeroInforme, c.catDescripcion, e.entNombre, u.uniNombre,
[Control].[fnDiasLaboralesFeriadosNacionalesLocales](o.ordtraFechaRegistro,i.infFechaAprobacion) TotalDias,
case
    when o.tipexaId='S' then 'Auditoría de Gestión'
    when o.tipexaId='B' then 'Auditoría de Obras Públicas'
    when o.tipexaId='F' then 'Auditoría 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 'Auditoría de Gestión'
end tipoExamen
from Control.OrdenTrabajo o
inner join Control.Informe i on o.ordtraId = i.ordtraId
inner join Catastro.Entidad e on e.entCodigo=o.ordtraEntidadExaminada
inner join General.catalogo c on c.catCodigo = e.entSector
inner join Talento.Unidad  u on u.uniCodigo=e.entAmbito
where year(ordtrafechaRegistro)=2022
and year(infFechaAprobacion)=2022
and i.infTipoInforme ='G'
```

Results | Messages

| | infNumeroInforme | catDescripcion | entNombre | uniNombre | TotalDias | tipoExamen |
|---|---|---|---|---|---|---|
| 1 | DNA1-0086-2022 | Administrativo | SECRETARÍA TÉCNICA DE GESTIÓN INMOBILIARIA DEL S... | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 160 | Examen Especial |
| 2 | DNA1-0078-2022 | Administrativo | SECRETARÍA TÉCNICA DE GESTIÓN INMOBILIARIA DEL S... | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 150 | Examen Especial |
| 3 | DNA1-0072-2022 | Administrativo | PRESIDENCIA DE LA REPÚBLICA | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 144 | Examen Especial |
| 4 | DPM-0025-2022 | Administrativo | SECRETARÍA NACIONAL DE PLANIFICACIÓN | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 167 | Examen Especial |
| 5 | DNA1-0092-2022 | Administrativo | DIRECCIÓN GENERAL DE REGISTRO CIVIL, IDENTIFICACI... | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 178 | Examen Especial |
| 6 | DNA1-0087-2022 | Administrativo | PROCURADURÍA GENERAL DEL ESTADO | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 170 | Examen Especial |
| 7 | DNA1-0093-2022 | Administrativo | SERVICIO NACIONAL DE CONTRATACIÓN PÚBLICA | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 178 | Examen Especial |
| 8 | DNA1-0062-2022 | Administrativo | SERVICIO ECUATORIANO DE NORMALIZACIÓN | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 121 | Examen Especial |
| 9 | DNA1-0069-2022 | Administrativo | INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC) | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 145 | Examen Especial |
| 10 | DNA1-0097-2022 | Administrativo | INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC) | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 93 | Examen Especial |
| 11 | DNA1-0083-2022 | Administrativo | INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC) | DNA 1 - DIRECCIÓN NACIONAL DE AUDITORÍA DE ADMIN... | 160 | Examen Especial |

## *Detalle de 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. |

# Desarrollo

## Unidades de Control

- DNA1 Administración Central
- DNA2 Sectores Sociales
- DNA3 Deuda Pública y Finanzas
- DNA4 Telecomunicaciones, conectividad y sectores productivos
- DNA5 - Gobiernos Autónomos Descentralizados
- DNA6 - Recursos Naturales
- DNA7 - Salud y Seguridad Social
- DNA8 - Transporte, Vialidad, Infraestructura Portuaria y Aeroportuaria

## Unidades y ámbitos seleccionados para el estudio

| Unidades | Ámbitos de control |
|---|---|
| **DNA1 Administración Central** | Administrativo |
| **DNA2 Sectores Sociales** | Bienestar Social |
| **DNA7 - Salud y Seguridad Social** | Salud |

# Desarrollo

**Tipo de exámenes de auditoria**

- **Aspectos Ambientales**
- **Obras Públicas**
- **Financiera**
- **Control Interno**
- **Examen Especial**
- **Auditoría de gestión**

**Tipo de informes**

- **General**
- **Penal**

# Desarrollo

## Indicadores de control

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

# Preprocesamiento

# Desarrollo

## Preprocesamiento de datos

- De los datos obtenidos y depositados en formato xlsx, se los analizó campo por campo y se va seleccionando los que aportan valor según los objetivos planteados. Mediante el preprocesamiento se identifican los datos, de forma que se retiran valores nulos, incorrectos, no válidos, desconocidos, entre otros.

- A los datos del archivo digital, luego de un estudio y análisis 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 exámenes cancelados, y se seleccionan los campos requeridos para el estudio.

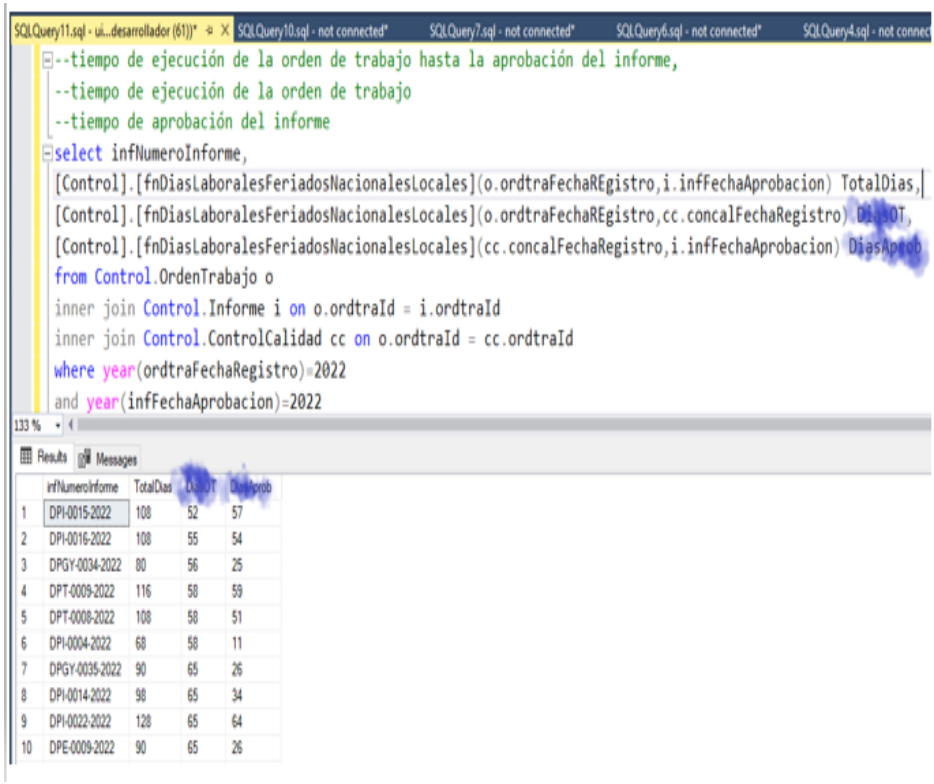
# Tansformación

# Desarrollo

## Transformación de datos

Se identifican 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, y se generan campos adicionales necesarios 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.

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.



```
--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
```

Results | Messages

| | 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 |

# Minería de Datos

# Desarrollo

## Técnicas seleccionadas para aprendizaje automático

**Aprendizaje Supervisado**

**Clasificación**

- Arboles de decisión
- Naive Bayes
- MultiClass Classifier
- JRIP

**Aprendizaje no Supervisado**

**Clustering o agrupación**

- K-means

# Herramientas para minería de datos

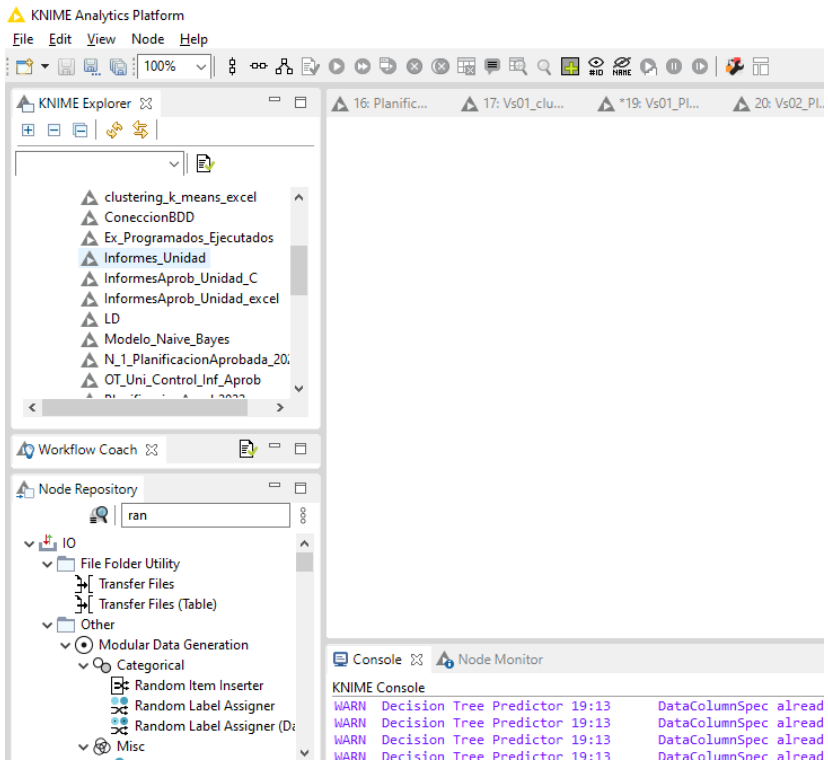
| Herramienta | Características | Lenguaje de programación | Sistema Operativo | Tipo de licencia |
|---|---|---|---|---|
| **RapidMiner** | Apto para todos los procesos. Destaca en el análisis predictivo | Java | Windows, macOS, Linux | Freeware, diferentes versiones de pago |
| **WEKA** | Muchos métodos de clasificación | Java | Windows, macOS, Linux | Software libre (GPL) |
| **Orange** | Crea una visualización de datos atractiva sin que se requieran muchos conocimientos previos para ello | Núcleo del software: C++, ampliación y lenguaje de entrada: Python | Windows, macOS, Linux | Software libre (GPL) |
| **KNIME** | Software de data mining de código abierto que ha democratizado el acceso a los análisis predictivos | Java | Windows, macOS, Linux | Software libre (GPL) (a partir de la ver.2.1) |
| **SAS** | Caro, pero potente para grandes empresas | Lenguaje SAS | Windows, macOS, Linux | Freeware limitado a instituciones públicas, el precio se establece tras solicitud, diferentes modelos disponibles |

# Desarrollo

## Minería de datos

Se hace uso de los algoritmos que posee la herramienta Knime se construyen los modelos con los siguientes algoritmos de minería de datos:

- Árboles de decisión
- Naive Bayes
- K-Means



# Desarrollo

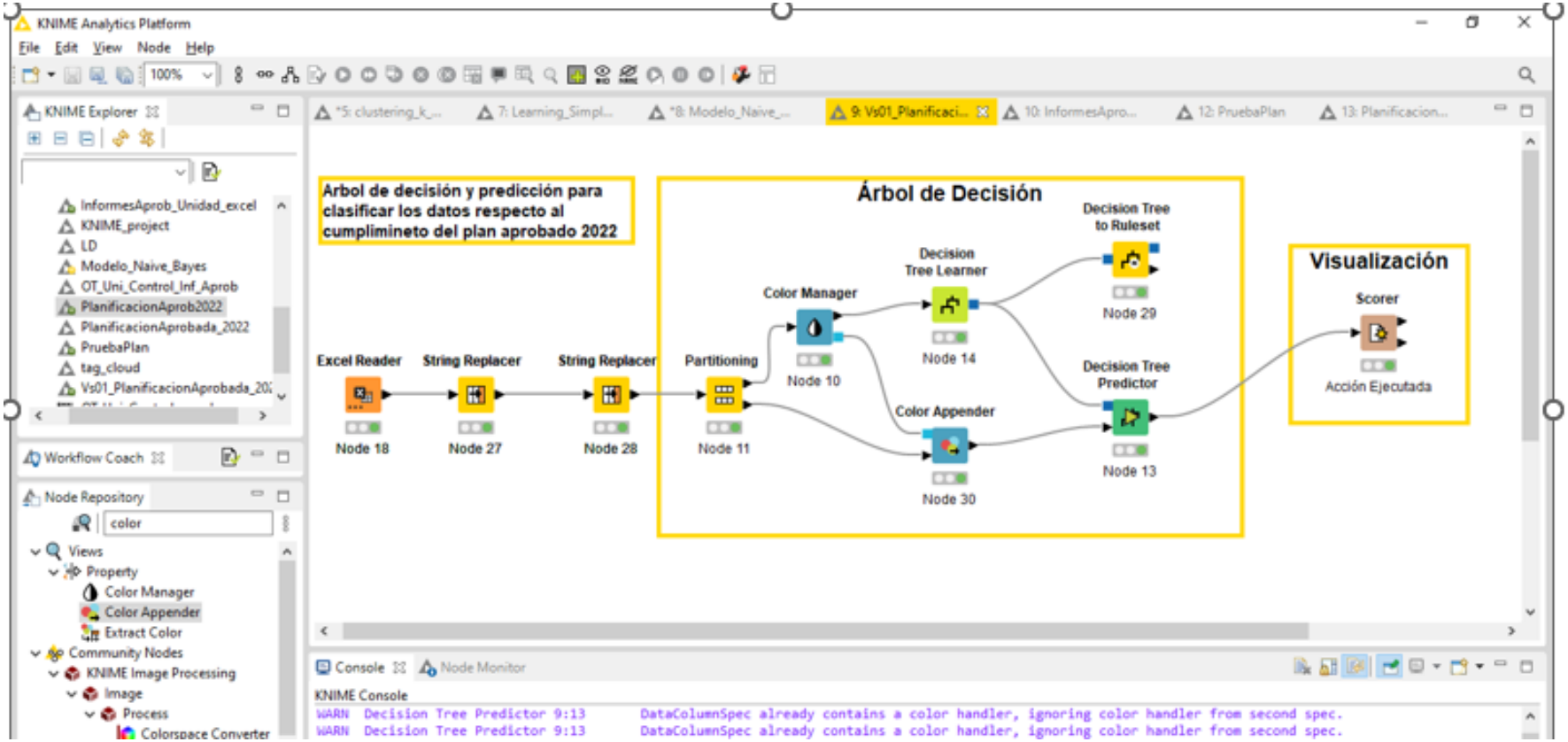
## Minería de datos

Se hace uso de los algoritmos que posee la herramienta Weka se construyen los modelos con las siguientes algoritmos de minería de datos:

- Naive Bayes
- MultiClass Classifier
- JRIP

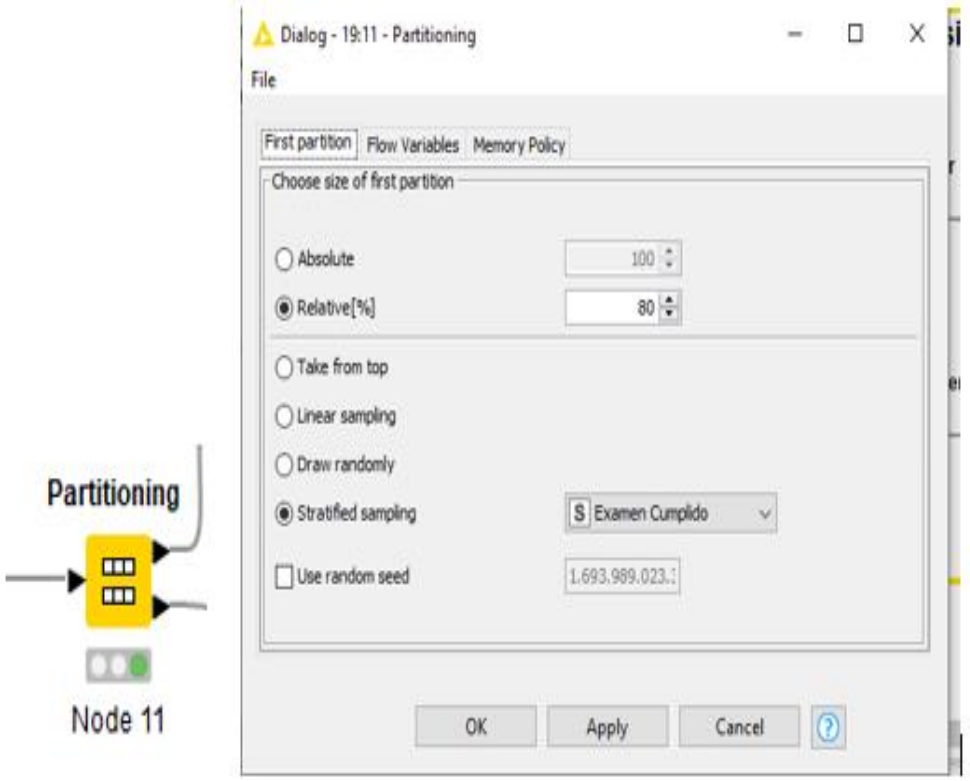
# Implementación del Modelo

*Modelo de Clasificación de Datos con árboles de decisión en Knime*



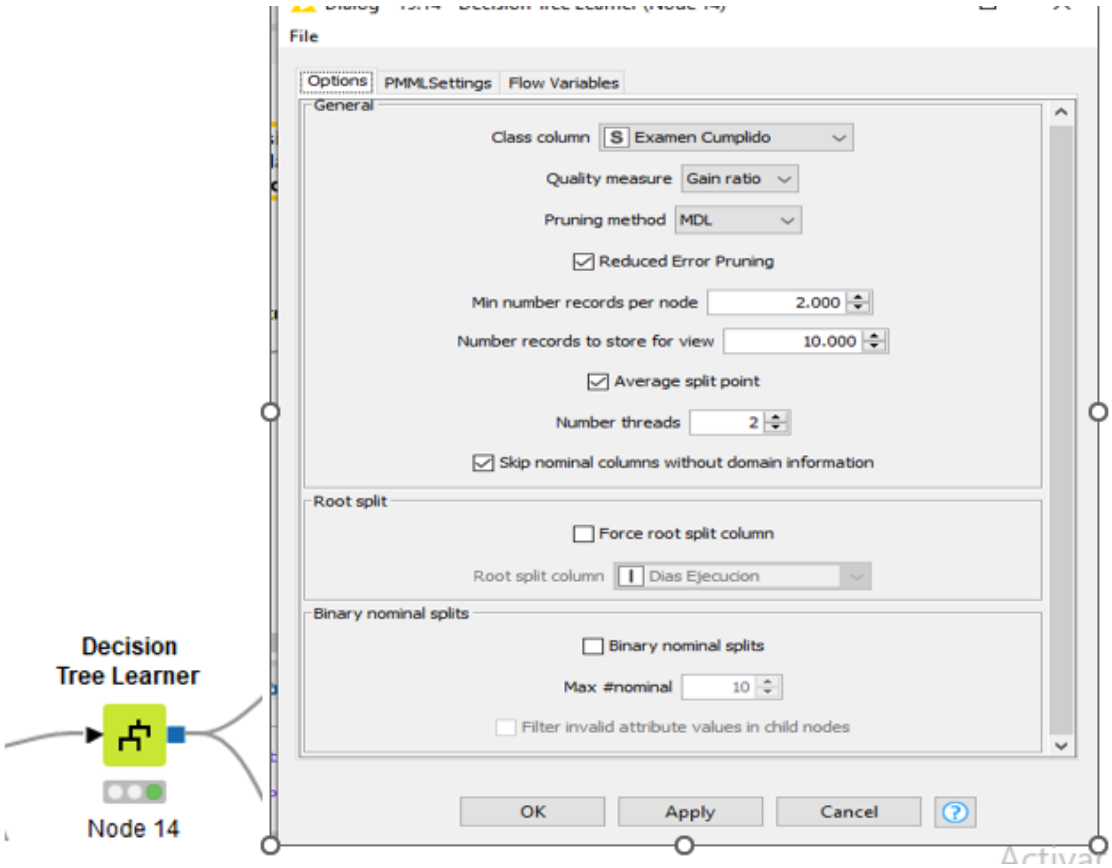
# Implementación del Modelo

*Configuración de nodo Partitioning del Modelo de árboles de decisión en Knime*



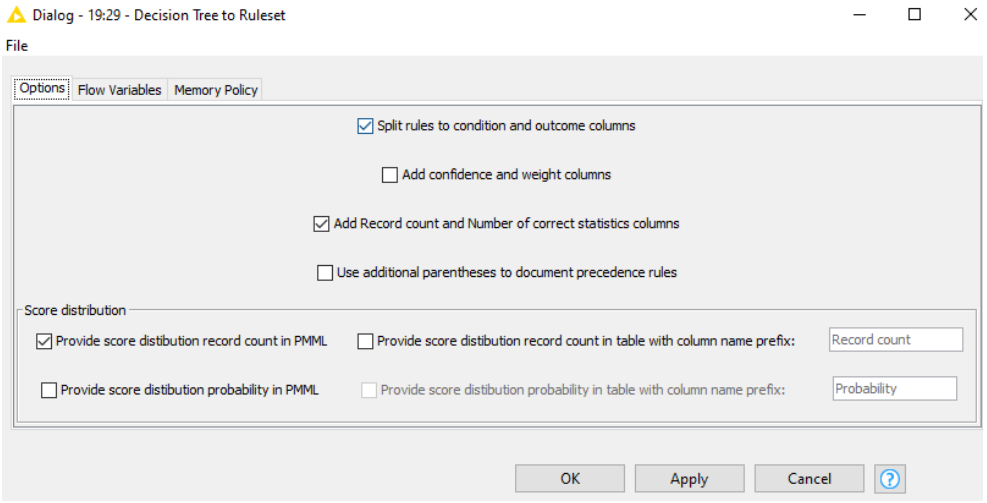
# Implementación del Modelo

*Configuración de nodo Decision Tree Learner del Modelo de árboles de decisión en Knime*



# Implementación del Modelo

*Configuración de nodo Decision Tree to Rulest del Modelo de árboles de decisión en Knime*


Rules table - 19:29 - Decision Tree to Ruleset

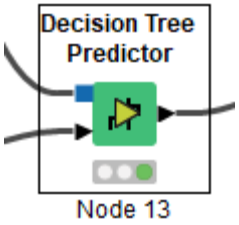
File Edit Hilite Navigation View

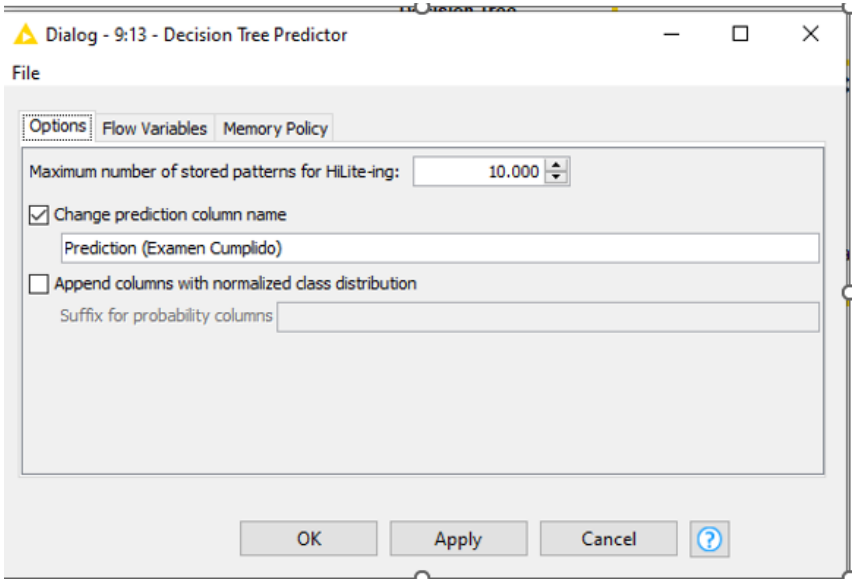
Table "default" - Rows: 2 | Spec - Columns: 4 | Properties | Flow Variables

| Row ID | S Condition | S Outcome | D Record ... | D Number... |
|---|---|---|---|---|
| Row1 | $Dias Ejecucion$ <= 31.5 AND TRUE | 0 | 304 | 304 |
| Row2 | $Dias Ejecucion$ > 31.5 AND TRUE | 1 | 65 | 65 |

# Implementación del Modelo

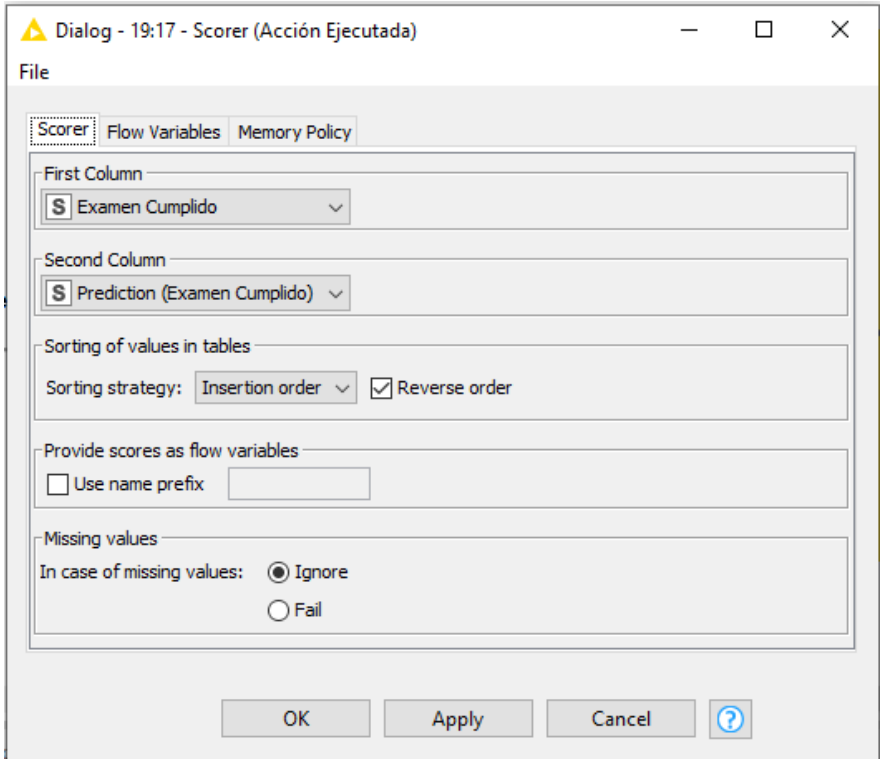
*Configuración de nodo Decision Tree to Rulest del Modelo de árboles de decisión en Knime*



# Implementación del Modelo

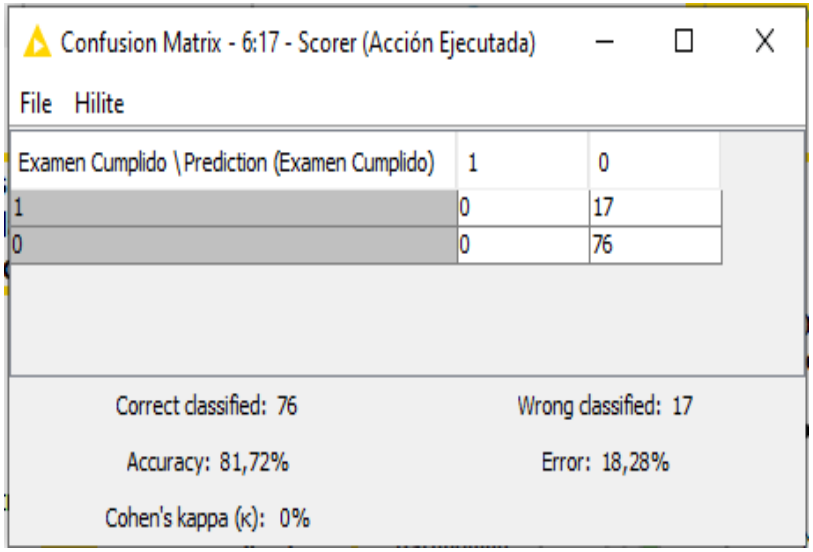
*Configuración de nodo Visualizacion del Modelo de árboles de decisión en Knime*

# Resultados

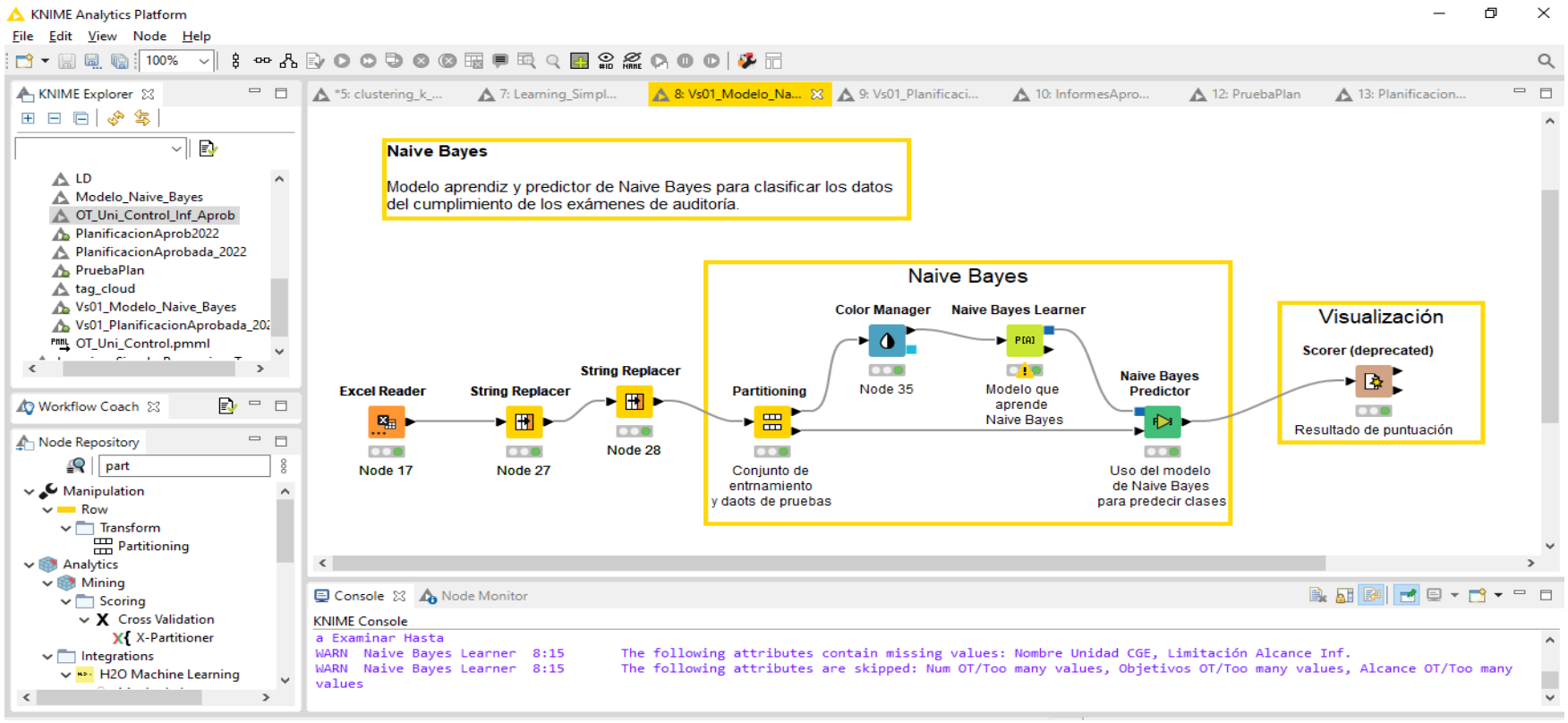
- ***Matriz de Confusión (Nodo Scorer) de Árboles de Decisión***

Nodo que se utiliza para la visualización de la salida de los resultados en la “Matriz de Confusión” con 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. Para el estudio se muestra si una acción de control se ejecuta.



# Implementación del Modelo

*Modelo de Clasificación de Datos de Naive Bayes en Knime*

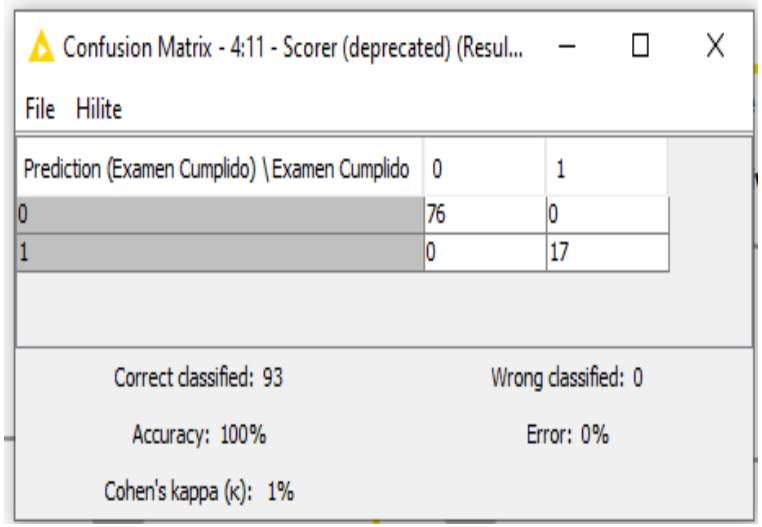


# Resultados

- ***Matriz de Confusión (Nodo Scorer) de Naive Bayes***

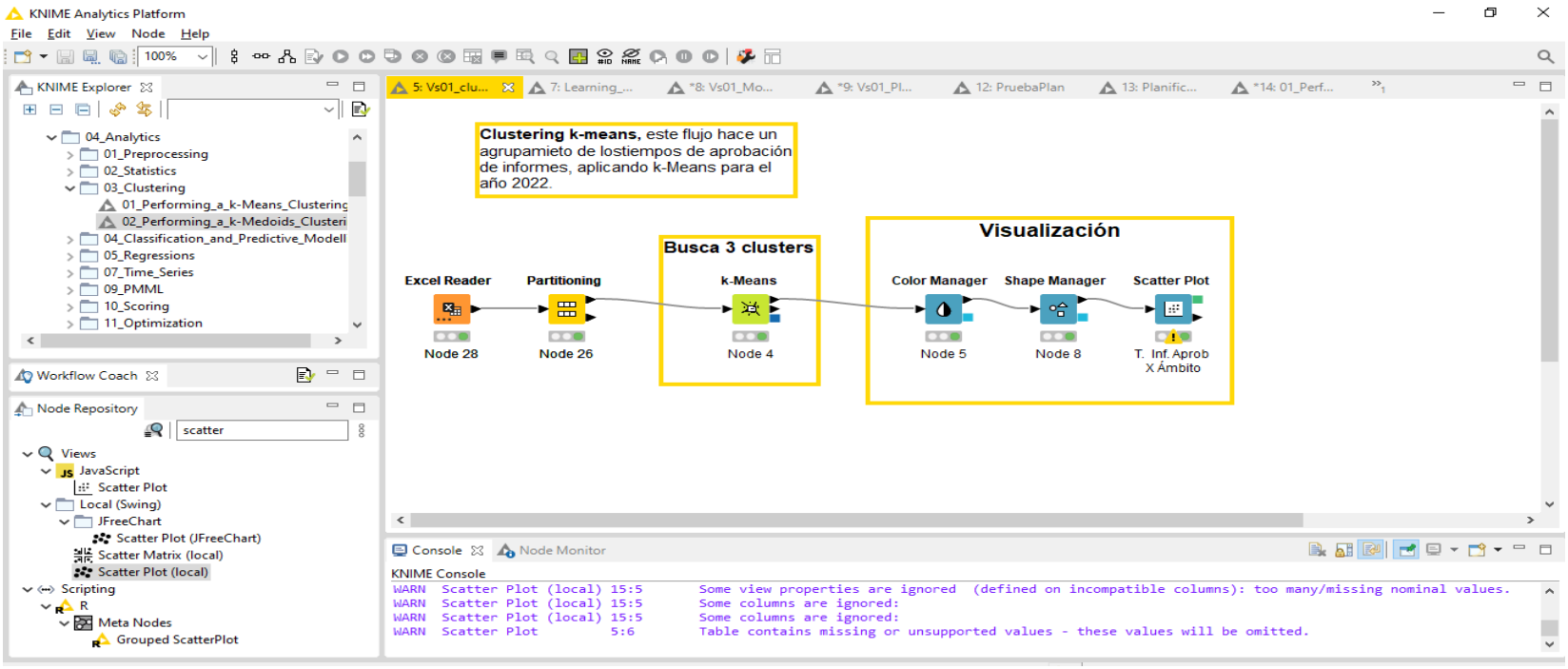
Permite visualizar los resultados de la matriz de confusión, del número de coincidencias en cada celda al comparar el campo de entrenamiento con el campo predicho, resultado que permite evaluar que tan bueno es el modelo. Para el estudio.

Muestra el resultado al evaluar 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.



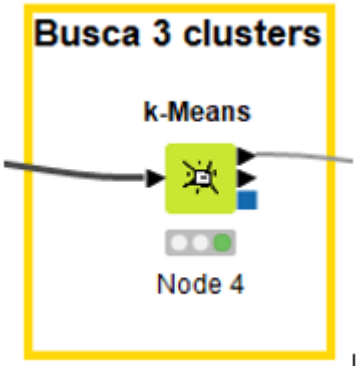
# Implementación del Modelo

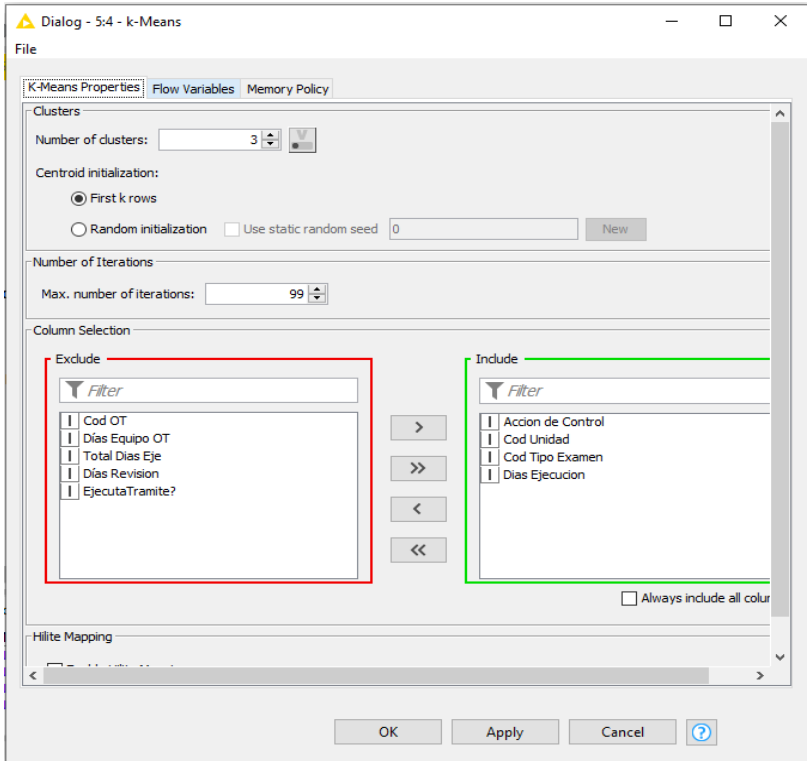
*Modelo K-means en Knime*



# Implementación del Modelo

*Configuración de nodo K-means del Modelo K-Means en Knime*



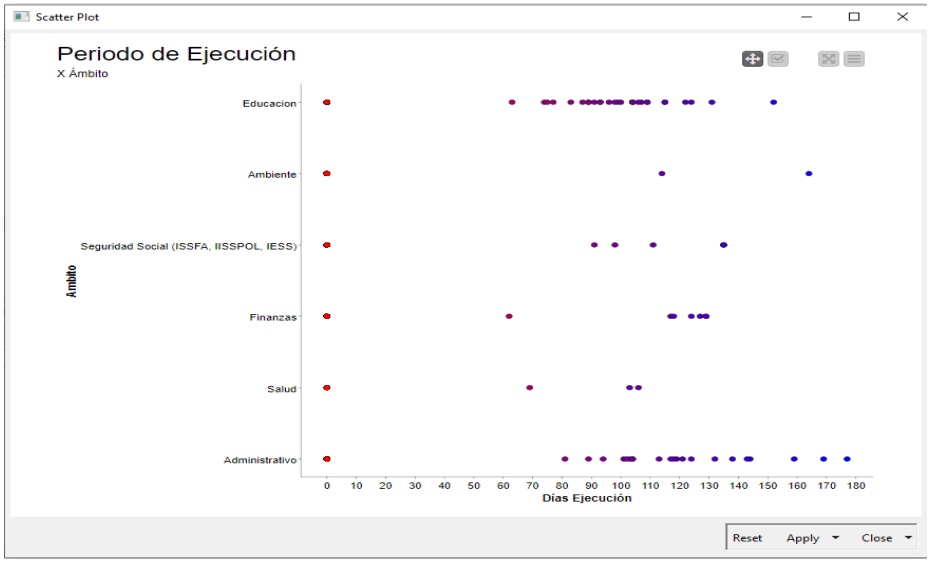



# Resultados

- ***Gráfico de Dispersión (Nodo Scatter Plot) modelo k-means***

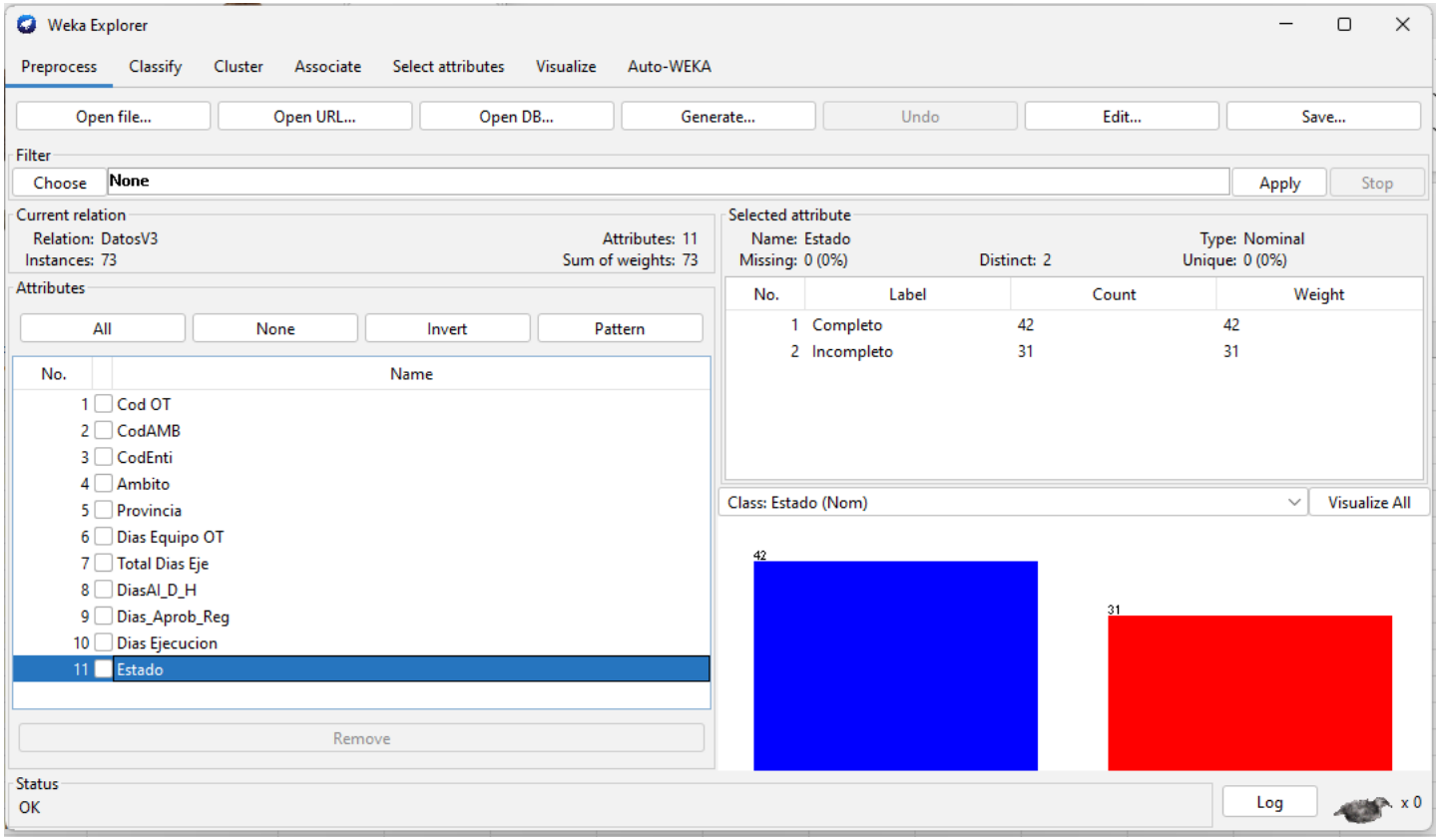
El modelo de K-Means nos permite evaluar los resultados a través de un gráfico de dispersión donde se ubican la mayor concentración de los puntos de un plano cartesiano, mostrando la correlación, pudiendo ser la correlación fuerte, débil o ninguna.

La concentración se da entre 80 y 140 días



# Resultados

*Interface de carga de datos y análisis con Weka*



# Resultados

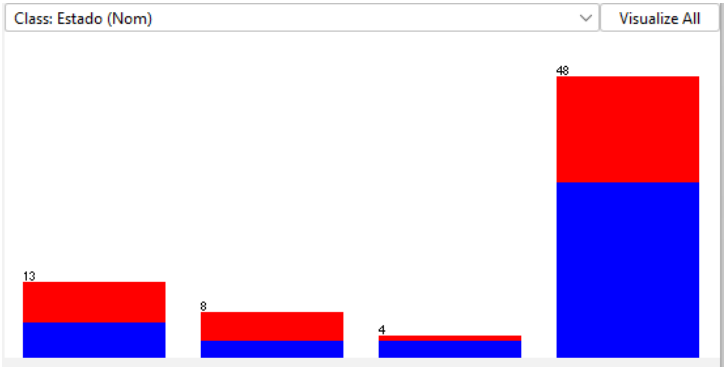
***Ejecución de la Orden de Trabajo según el ámbito con Weka***

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

Selected attribute

Name: Ambito Type: Nominal
Missing: 0 (0%) Distinct: 4 Unique: 0 (0%)

| No. | Label | Count | Weight |
|---|---|---|---|
| 1 | Finanzas | 13 | 13 |
| 2 | Ambiente | 8 | 8 |
| 3 | Salud | 4 | 4 |
| 4 | Educacion | 48 | 48 |



# Resultados

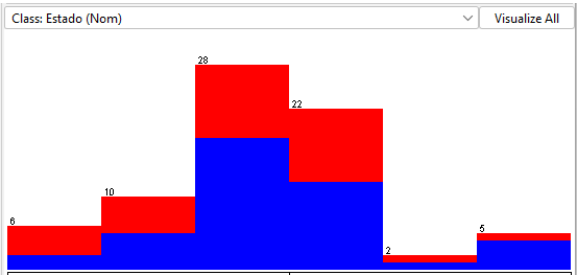
*Cantidad de Días de Ejecución de la Auditoría de la Orden de Trabajo* ***con Weka***

- 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

Selected attribute

Name: Dias Equipo OT  Type: Numeric
Missing: 0 (0%)  Distinct: 22  Unique: 7 (10%)

| Statistic | Value |
|---|---|
| Minimum | 40 |
| Maximum | 90 |
| Mean | 63.589 |
| StdDev | 10.789 |



# Resultados

***Matriz de Confusión del Modelo de Arboles de Decisión con Weka***

```
=== 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
```

# Resultados

## *Matriz de Confusión del Modelo de MultiClassClassifier*

```
=== 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 48 |  d = Educacion
```

## *Matriz de Confusión del Modelo de Arboles de Decisión con Weka*

```
=== 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
```

# Analisis de Resultados

## Interpretación y Evaluación

- Para el presente estudio se propuso varios modelos de clasificación y se ejecutaron en dos herramienta Knime y Weka.

- Arboles de decisión, Naive Bayes, JRip y MultiClass Classifier, ZeroR. La comparación de los 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.

# Analisis de Resultados

- Hacemos una comparativa de resultados de las Matrices de Confusión generadas en los Modelos de Clasificación con Knime

- 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.

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

# Analisis de Resultados

- Se hace una Comparativa de Resultados de las Matrices de Confusión generadas en los Modelos de Clasificación con Weka con otros campos y con otro banco de datos
- Se observa que el modelo de clasificación multiclase empleado en el aprendizaje automático y la clasificación estadística es el que mejor se adapta para el conjunto de datos usado.

| Técnica | Correctamente clasificados | Exactitud | Coeficiente 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 |

# Análisis de Resultados

## Interpretación y Evaluación

- Se realizó la evaluación de resultados con otros campos y otro banco de datos para los modelos de predicción y clasificación. Los 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 que el modelo de Naive Bayes presenta mejores resultados para el presente estudio.

- se observa que para el modelo Naive Bayes existe una buena correlación de los datos, por ejemplo, la correlación entre el examen ejecutado y el tiempo de cumplimiento, este modelo alcanzó una mayor exactitud.

# Conclusiones

- Se realizó la revisión de literatura para determinar las técnicas predictivas. Esto permitió seleccionar los indicadores sensibles en la operación auditada de la entidad.

- Se ha logrado estudiar y ejecutar varios modelos analítico-predictivo en 2 herramientas de análisis de datos con técnicas de minería de datos y observar el comportamiento de dichos modelos.

# Conclusiones

- Se han realizado diferentes pruebas y 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 en la Institución.

- Se realizó un estudio amplio del proceso de auditoría 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 puede publicar.

# Conclusiones

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 la 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 de ciertos campos no permitir el guardado con campos vacíos.

- El proceso de la planificación 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.

# Agradecimientos

A mi Tutora, Ing. Sonia Cardenas, Ph.D., quien con su experiencia supo guiarme, fue de gran apoyo en todo momento y con su entereza logramos alcanzar el objetivo juntas.

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

GRACIAS