

# Analítica de Datos Sociales

Jose Aguilar

Febrero 2017

**Profesor:** Jose Aguilar

**Sitio de Trabajo:** Prometeo EPN, Yachay EP y  
UTPL

**Contacto:** aguilar@ula.ve

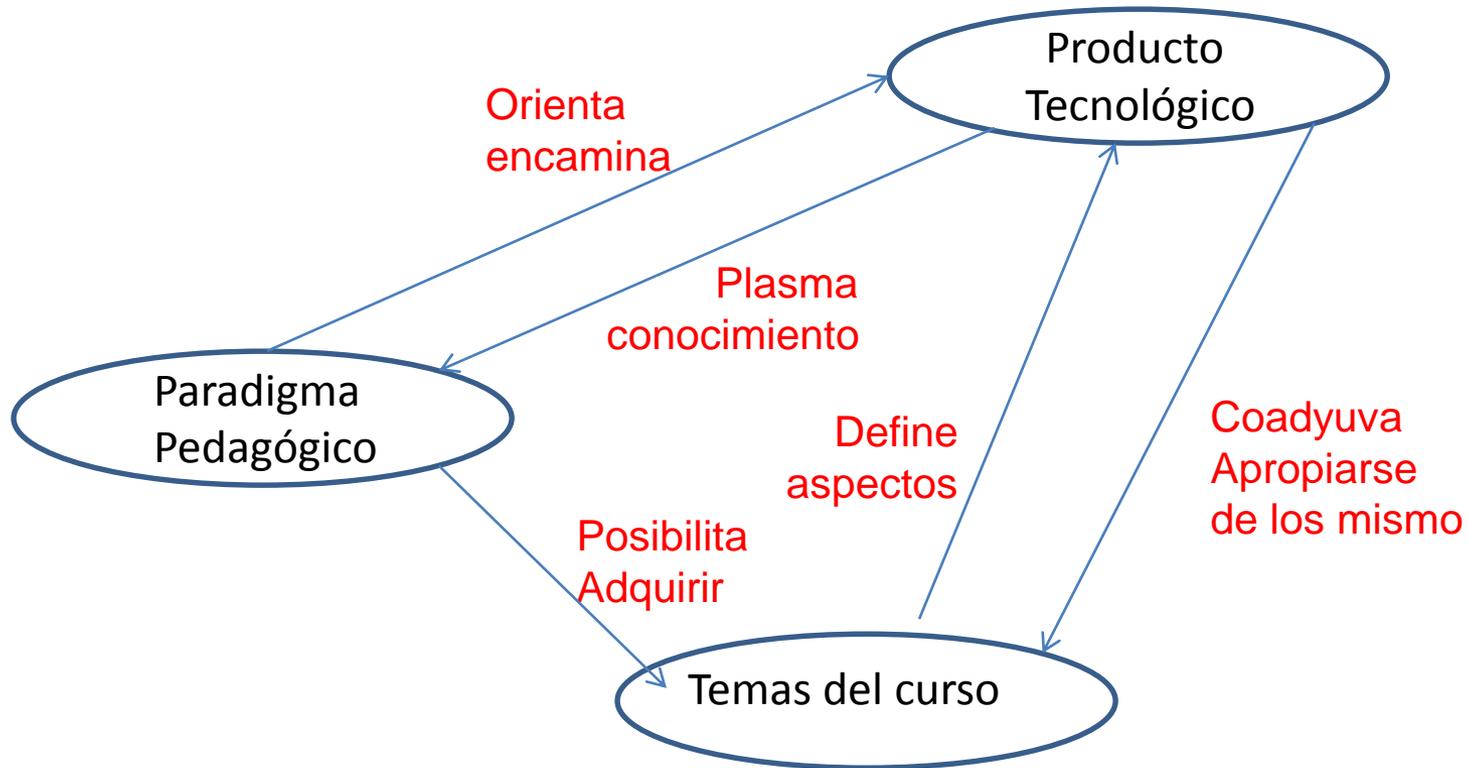
**Consulta:** por email cuadrar cita

**Información del curso:** lista, dropbox y  
[www.ing.ula.ve/~aguilar](http://www.ing.ula.ve/~aguilar)

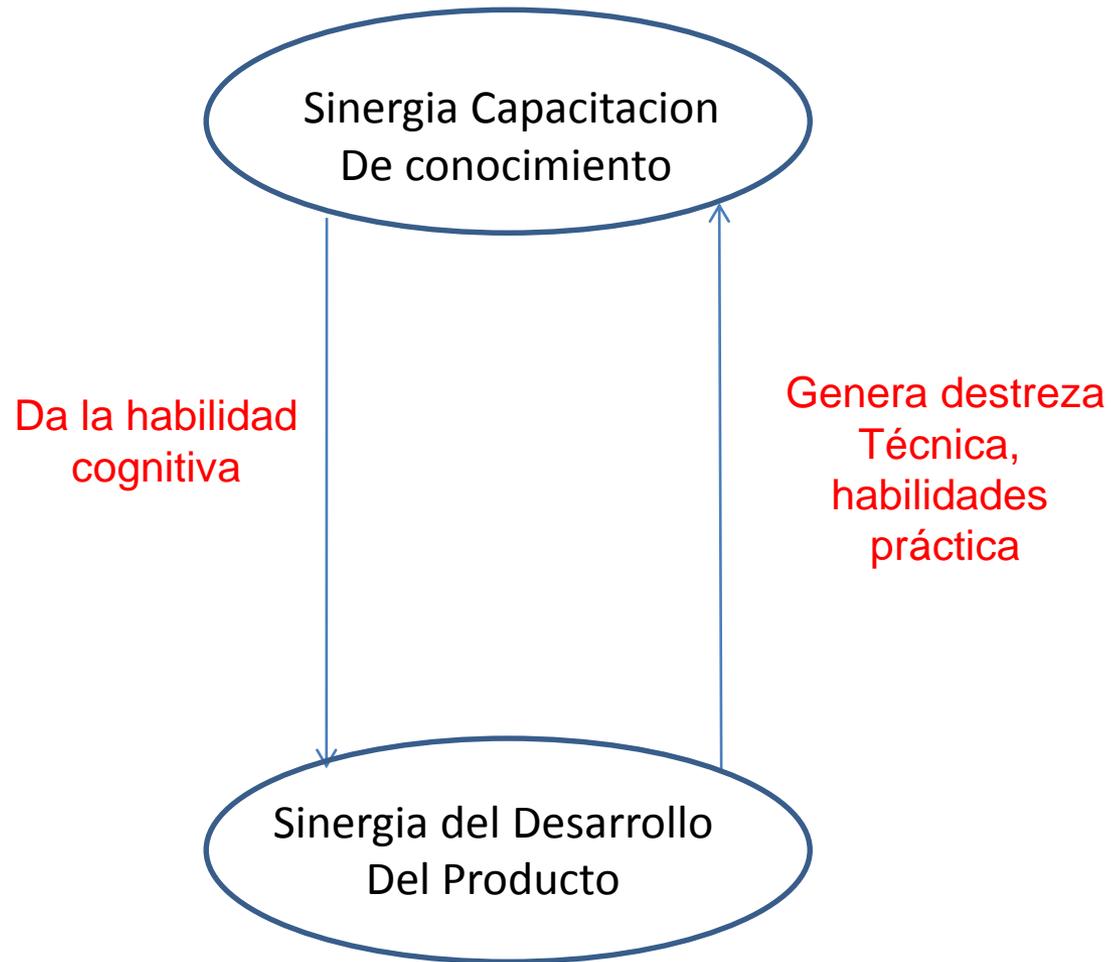
# OBJETIVO

Introducir a los participantes en la *Analítica de Datos Sociales*. En específico, los conceptos de base, metodologías para desarrollar tareas de analítica de datos social, conceptos vecinos, herramientas, etc.

# Dinámicas del Curso



# Dinámicas del Curso



# Dinámicas Sinergias

- Nos moveremos en las dos esferas
- Se introducirán los conceptos de base

# Sinergias

## Sinergia Desarrollo de Productos (SDP)

- Se trabajará en obras/proyectos,
- Trabajarán en grupos
- Los productos (obras) irán plasmando el conocimiento adquirido
- Los productos (obras) al final deberán contener todo el contenido adquirido en el curso inmerso en ellos
- Además del trabajo en clase se requiere también de trabajo fuera de las clases.
- ¿Cómo se trabaja?:
  - Metodología Ágil para la gestión y construcción del proyecto
  - Se debe dar cuenta del recorrido del desarrollo del producto diariamente (dejan informes de avances en dropbox)

# Contenido

Día	Actividad Docente	Actividad Práctica
13.02.17	Tema 1. Bases filosóficas del Curso y Metodologías de la Analítica de Datos a usar.	Definición del contexto del Problema
14.02.17	Tema 2. Introducción a la Analítica de Datos Sociales, SNA	Conocer el contexto del problema
15.02.17	Tema 3. Ciencia de los Datos para la Analítica de Datos Sociales Tema 4: Técnicas de Analítica de Datos Sociales: Minería semántica, Minería de texto, Minería de Grafos,.	Diseño del ciclo autonómico de tareas de Analítica de Datos Presentación Avances
16.02.17	Tema 4: Técnicas de Analítica de Datos Sociales: Minería de Grafos, enlazado de datos, etc.	Diseño de las tareas de Analítica de Datos Social
17.02.17	Tema 5. Conceptos Vecinos: Datos Masivos (BigData), Inteligencia de Negocios	Preparar Implementación y Experimentación Presentación Avances

# Recorrido Sinergia Desarrollo de Productos

- Se seguirá metodología **yPBL** , **MIDANO** y **aprendizaje semi-invertido**
  - **yPBL:**
    - Ramas de: diseño (funcional y técnico) y desarrollo
    - Iterativo (ágil)
    - Fases de análisis, diseño, implementación y pruebas
  - **MIDANO-ext**
    - Especificación del ciclo autonómica de Analítica de datos Social
    - Diseño de la tarea de Analítica de datos Social
  - **CRISP-DM**
    - Desarrollo de la tarea de Analítica de datos Social

# yPBL

- **Metodología de aprendizaje inspirada en Ingeniería de software**
- **Permite construir aplicaciones reales de software mientras se aprende.**
- **Cada Iteración:**
  - Cubre un tópico del curso aplicado al producto tecnológico
  - Se redefinen roles en los grupos, recursos usados, cronogramas
  - Interactuamos todos para alcanzar los objetivos de aprendizaje

# Desarrollo del curso

## Requerimientos:

Lista de requerimientos  
(Por iteración)

## Análisis:

Entender el problema y plantear  
posibles soluciones

## Diseño:

Proponer una solución que satisfaga  
parcial o totalmente los requerimientos

## Implementación:

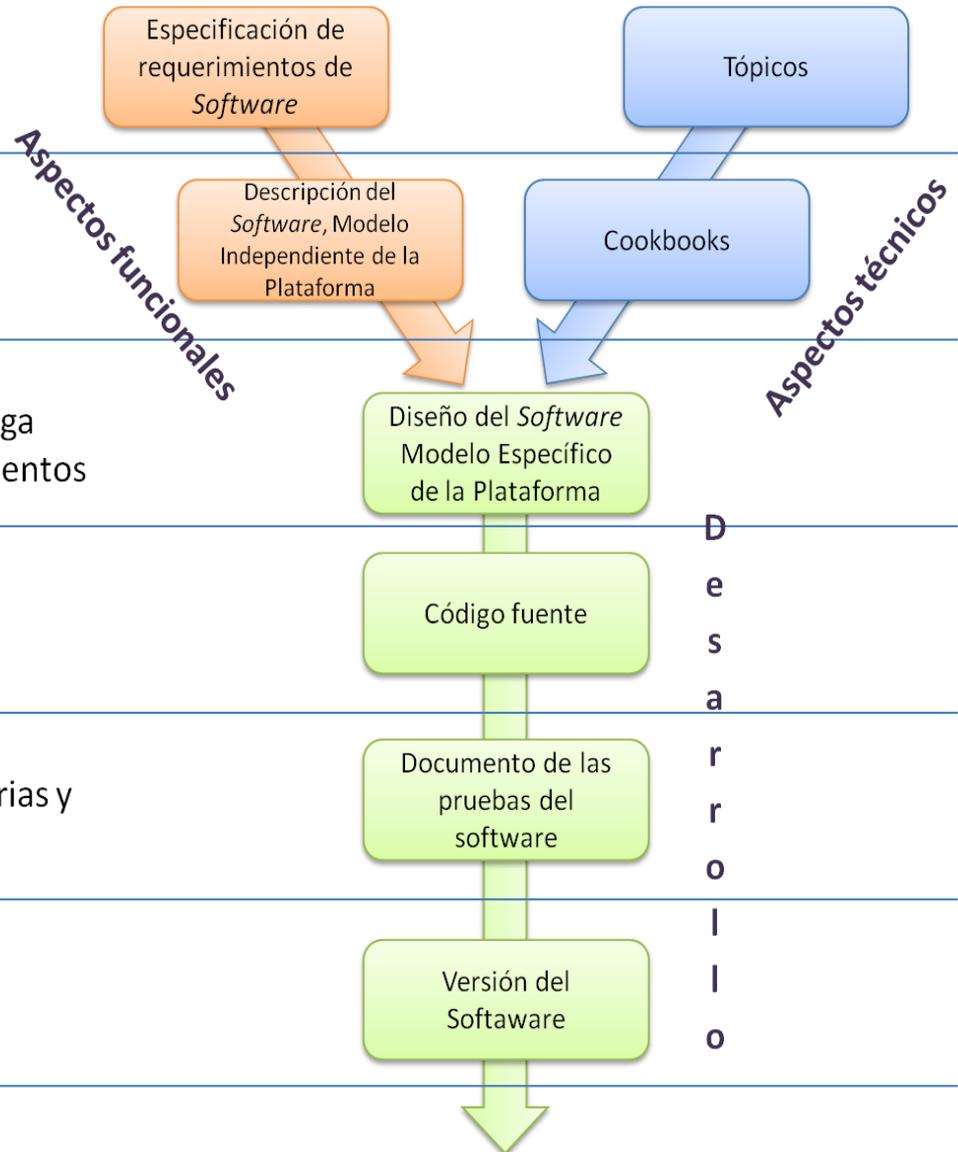
Desarrollar la solución

## Prueba:

Evaluar la solución (pruebas unitarias y  
de integración)

## Entrega:

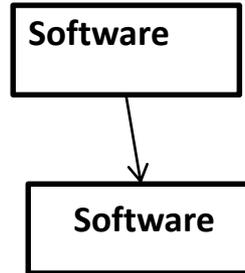
Entregar la solución al cliente



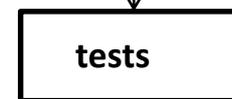
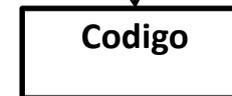
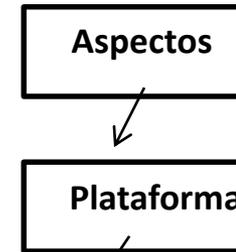
# yPBL

- **Requerimientos**
- **Análisis**
- **Diseño**
- **Implementación**
- **Pruebas**
- **Liberación**

## Rama Funcional



## Rama Técnica



## Rama Desarrollo

## Iteraciones

I1

Conocer el contexto del problema

I2

Diseño del ciclo autónomo de tareas de Analítica de Datos

I3

Diseño de las tareas de Analítica de Datos Social

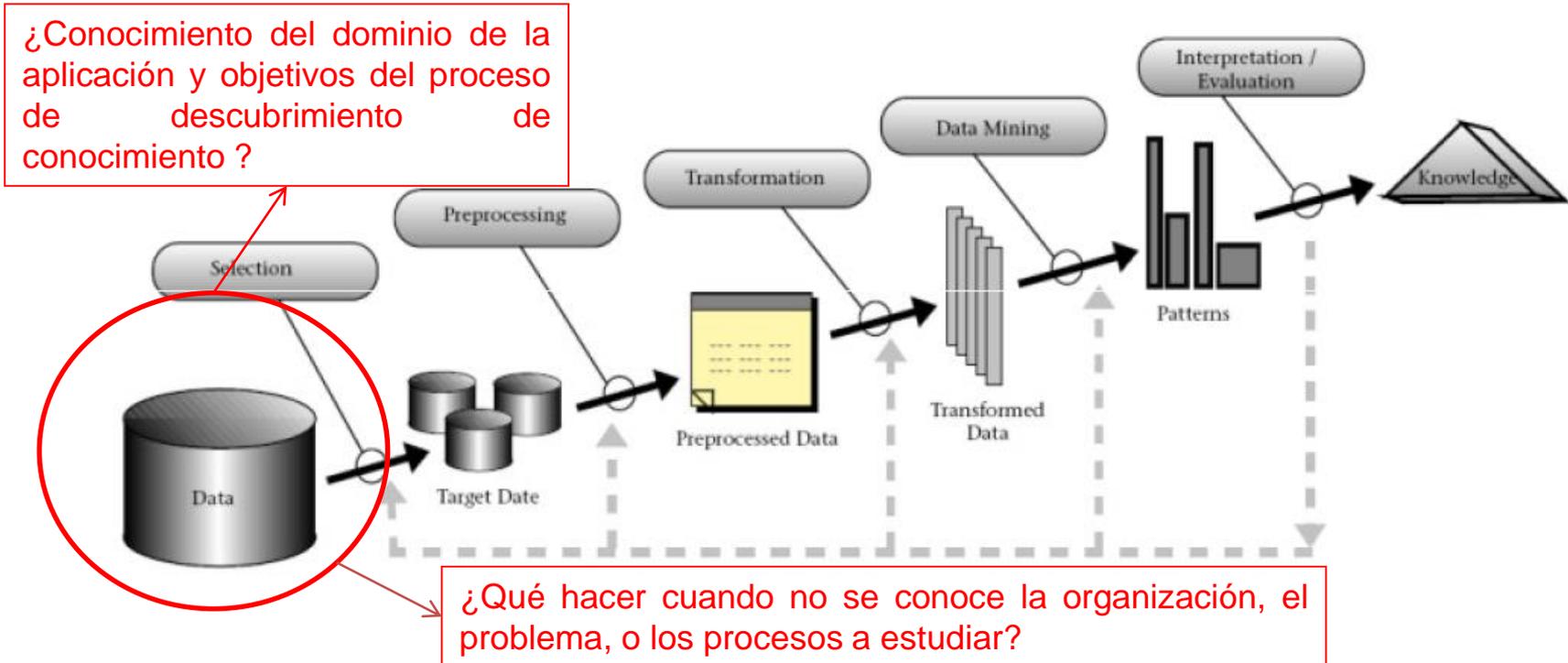
# Metodologías para realizar Analítica de Datos en una organización

# MIDANO

**“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”**

**Extendida para ser usado en el análisis de datos**

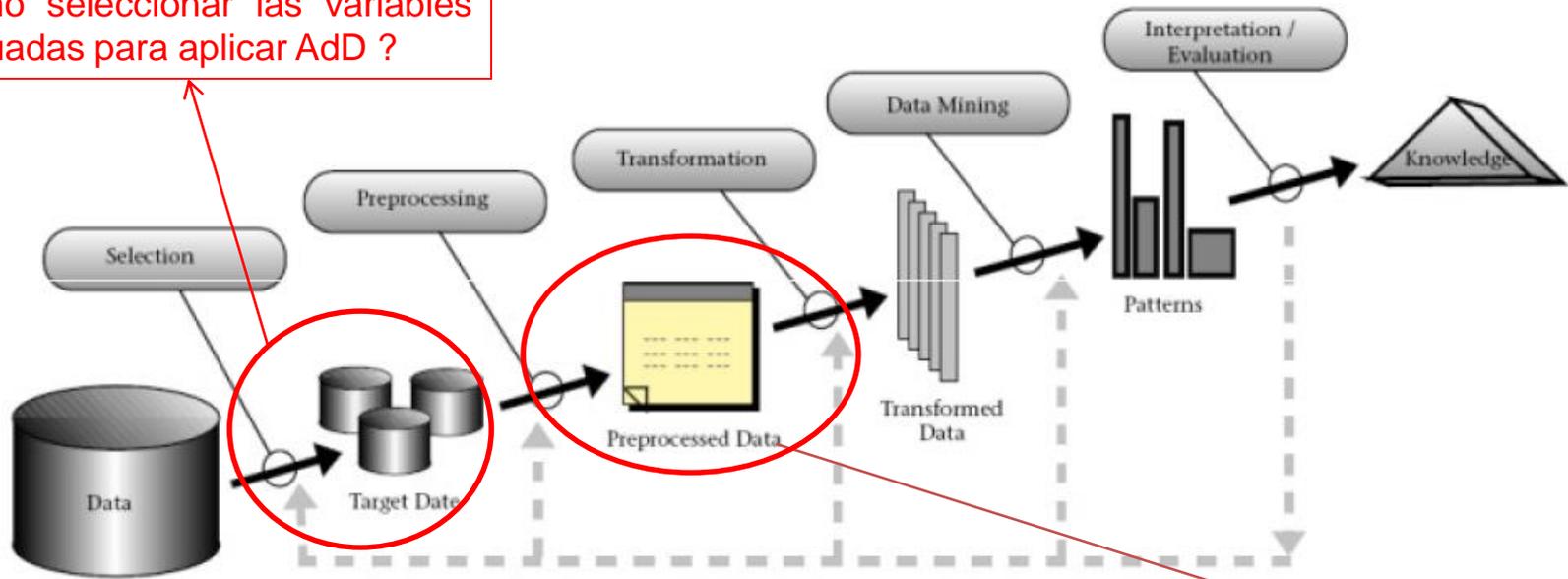
# MIDANO



**“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”**

# MIDANO

¿Cómo seleccionar las variables adecuadas para aplicar AdD ?



¿Cómo realizar el procesamiento de datos?

**“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”**

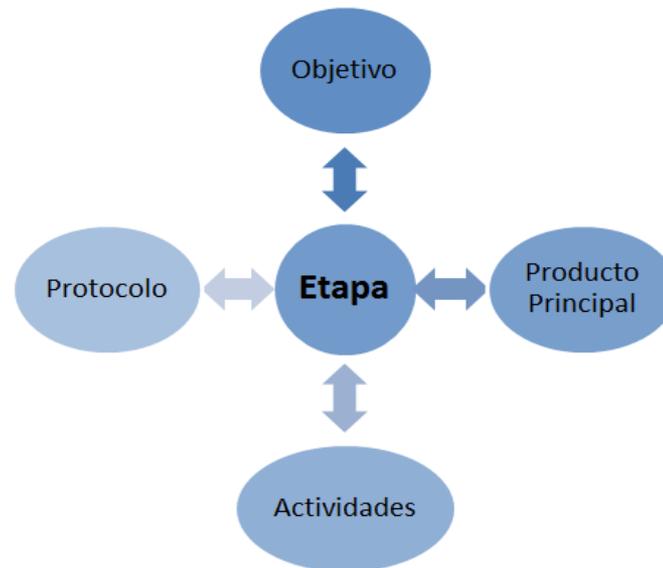
# MIDANO-AdD

MIDANO consta de tres fases.



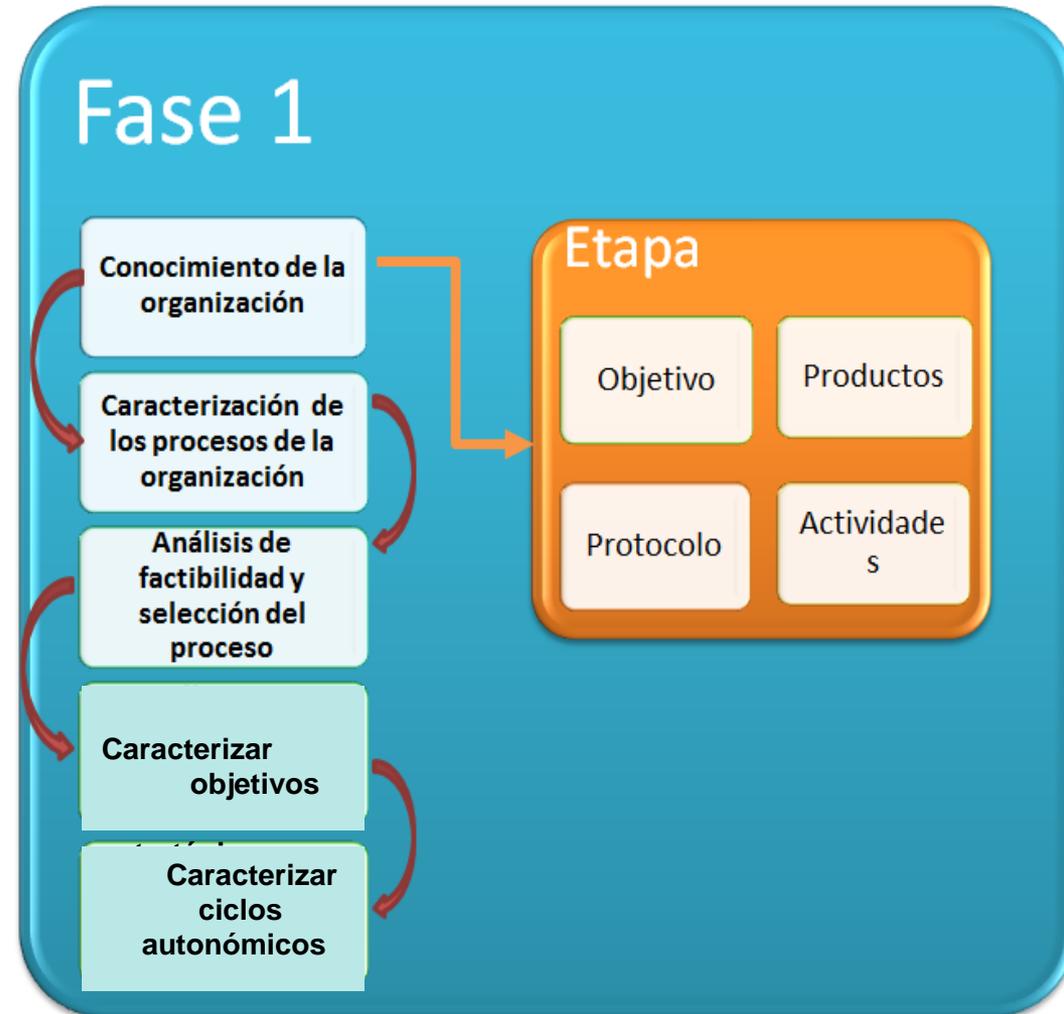
# MIDANO

Cada fase de la metodología está dividida en etapas, los elementos que describen cada etapa. En cada etapa se especifican cuatro aspectos principales, como se describe a continuación.



# Fase 1: Conocimiento de la Organización

Esta fase tiene como finalidad realizar un proceso de ingeniería de conocimiento, orientado a organizaciones/empresas, de las cuales no se conoce o se tiene poca información del (de los) problema(s), o los procesos a estudiar.



# Etapa 1: Conocimiento de la Organización

1. Objetivo {
- Conocer la organización/empresa, sus objetivos, procesos, objetos y actores

## 2. Protocolo de la Fase:

- Descripción de los elementos de la institución/empresa y sus características. Objetivos, Procesos , Objetos y Actores.
- Descripción de las relaciones entre estos elementos.
- Organización de estos elementos.

# Etapa 1: Conocimiento de la Organización

Preguntas y ejemplos para determinar los elementos de la institución/empresa

Elemento	Preguntas	Ejemplos
Objetivos	¿Cuál es la razón de ser de la institución?	Conocer, determinar, establecer, la finalidad de la institución/empresa.
Procesos	¿Cuales son las actividades que permiten alcanzar los objetivos de la institución?	Procesos de producción o administrativos.
Objetos	¿Qué cosas o entidades se manipulan en los procesos de la institución?	Pueden ser físicos o abstractos, departamentos, documentos, herramientas, plantas.
Actores	¿Quiénes ejecutan los procesos?	Personas, sistemas, máquinas, etc.

## Etapa 2: Caracterización detallada de los procesos de la organización

1. Objetivo {
- Conocer los procesos sobre los cuales se puede enfocar el proyecto de AdD.

### 2. Protocolo de la Fase:

- Familiarización con los procesos sobre los cuales se puede realizar la ingeniería de conocimiento
- Identificación de la fuente de conocimiento
- Familiarización con los ambientes computacionales donde se encuentran los datos a ser utilizados en cada proceso.

# Etapa 2: Caracterización detallada de los procesos de la organización

## 1. Familiarización con los procesos sobre los cuales se puede realizar la extracción de conocimiento

- ¿Qué productos generan esos procesos?
- ¿Qué beneficios proporcionan esos procesos a la organización?
- ¿Qué problemas tienen actualmente?
- ¿Importancia de esos procesos para la organización, o impacto sobre otros procesos?
- ¿Qué impacto generaría la mejora de esos procesos o el estudio de los mismos?

## 2. Identificar la fuente del conocimiento

- ¿Cuáles son los actores o personas que intervienen en los procesos?
- ¿Quién o quiénes son las personas expertas en los procesos?
- ¿Existen documentos que permitan conocer esos procesos?
- ¿Existen sistemas computacionales que intervengan o interactúen en el proceso?

## 3. Familiarización con los ambientes computacionales donde se encuentran los datos a ser utilizados en cada proceso explicado

- ¿Dónde se encuentran los datos almacenados del proceso en cuestión?
- ¿Cómo se almacenan los datos del proceso?
- ¿Qué variables son observadas del proceso?
- ¿Cuáles son las variables más importantes de esos datos para la organización?

# Etapa 3: Análisis de factibilidad y selección de los procesos

1. Objetivo
- 
- Analizar los procesos con la información proporcionada/recogida, con la finalidad de conocer la factibilidad de la aplicación de la AdD sobre cada uno de ellos

## 2. Protocolo de la Fase:

- Revisión de los procesos propuestos por los expertos
- Disponibilidad del experto o grupo de expertos
- Análisis de las fuentes de información sobre los procesos

# Etapa 3: Selección de los Procesos

## Ejemplo de Tabla para selección de procesos

Peso	Criterios	Proceso 1	Proceso 2
	Importancia para la organización		
	Interacciones entre procesos		
	Procesos dependientes		
	Importancia de la calidad del producto		
	Seguridad Industrial		
	<b>Proposito de la tarea de Add</b>		
	Replicabilidad de la herramienta a desarrollar		
	Cantidad de Expertos		
	Fuentes de información		
	Confidencialidad de la información		
	¿Qué información se recoge del proceso para ser almacenada?		
	Con que frecuencia se recoge la información almacenada		
	¿Qué herramientas se cuentan, para recolectar y manipular la información?		
	Total sin ponderación		
	Total ponderado		

Criterios vinculados a la importancia del proceso para la organización

Criterios vinculados a la factibilidad de hacer una Tarea de Análítica de Datos

# Etapa 4: Análisis para caracterizar los objetivos estratégicos a alcanzar

## 1. Objetivo

- Caracterizar las posibles objetivos estratégicos a alcanzar, con las tareas de AdD, en los procesos seleccionados

## 2. Protocolo de la Fase:

- Descripción de los escenarios actuales de los procesos seleccionadas en la institución/empresa.
- Especificación de los objetivos estratégicos a alcanzar en esos procesos, y posibles escenarios futuros detrás de ellos.
- Especificación de los indicadores (modelos de conocimiento, medidas estadísticas, etc.) para el análisis e interpretación de los objetivos estratégicos
- Especificación de los requerimientos para los posibles escenarios futuros (donde se puedan aplicar tarea(s) de AdD)
- Elaboración de los casos de uso para los requerimientos funcionales

## ***Etapa 4: Análisis para caracterizar los objetivos estratégicos a alcanzar***

**Para los procesos seleccionados**

### **Descripción del escenario actual**

<b>Resultados que se obtienen</b>	<b>Actor(es) asociado(s)</b>	<b>Variables Asociadas</b>	<b>Actividades que se realizan</b>

# Etapa 4: Análisis para caracterizar los objetivos estratégicos a alcanzar

Para los procesos seleccionados:  
todos sus posibles escenarios futuros

Escenarios futuros deben estar orientados a lograrlos

Métricas estadísticas, modelos de conocimiento, ...

## Descripción del escenario futuro

Objetivos Estratégicos a alcanzar	Actor(es) asociado(s)	Variables Asociadas	Actividades de AdD que se realizarían	Funcionalidades nuevas	Resultados que se desean obtener (indicadores de logro)

Descripción del escenario futuro: < xxx >

El conjunto de escenarios futuros define una **planificación estratégica tecnológica organizacional**

# Etapa 4: Análisis para caracterizar los objetivos estratégicos a alcanzar

## Priorización de los escenarios futuros

Criterios	Escenario 1	Escenario 2	Escenario 3
Importancia del resultado que se espera del escenario para la empresa/institución			
Utilidad del escenario para la empresa/institución			
Cantidad de expertos asociados al escenario			
Seguridad Industrial (si aplica)			
Fuentes de información requeridas por el escenario			
Confidencialidad de la información			
¿Con que frecuencia se recogen los datos almacenados asociados a la información de interés?			
¿Con qué herramientas se cuenta para recolectar y manipular los datos?			
Replicabilidad de la herramienta a desarrollar en otros escenarios			

Vinculados a los **objetivos estratégicos** y su importancia

# Etapa 5: Caracterización de los ciclos autónomos de AdD para cada Objetivo Estratégico

## 1. Objetivo

- Especificación de los Ciclos Autónomos (CA) para cada escenario futuro (objetivo estratégico) priorizado

## 2. Protocolo de la Fase:

- Determinación de las tareas de AdD que deben caracterizar a c/ciclo por sus roles
  - Tareas de monitoreo
  - Tareas de análisis
  - Tareas de toma de decisión
- Especificación de las relaciones entre ellas
- Especificación general de las fuentes de datos requeridas por cada tarea

# Especificación del Ciclo Autónomo

**Objetivo:** Definir un objetivo válido de supremo interés para el proceso a estudiar.

## **Procedimiento General**

**Paso 1 Tareas de Monitoreo:** Se identifican, capturan, pre-procesan, las variables del proceso bajo estudio, para poder tener una **observación** clara del proceso bajo estudio

**Paso 2: Tareas de análisis:** Se **interpretan** las situaciones que va aconteciendo en el proceso que se está estudiando, para comprenderlo, diagnosticarlo, analizarlo, entre otras cosas.

**Paso 3 : Toma de decisiones:** Se definen **acciones a tomar** sobre el proceso, con el fin de alcanzar el objetivo definido para el ciclo.

# Etapa 5: Caracterización de los ciclos autónomos de AdD para cada Objetivo Estratégico

## Por cada ciclo autónomo

Objetivo estratégico a alcanzar: < ... >

	Nombre	Fuentes generales de datos requeridas	Indicadores generados	Efectos esperados sobre el objetivo estratégico
Tareas de AdD de Observación				
Tareas de AdD de Análisis				
Tareas de AdD de Toma de decisión				

Métricas estadísticas, modelos de conocimiento, ... que produce

Usado en el futuro como métrica de calidad del CA

## Relaciones entre las tareas del CA de AdD

	Tarea AdD1	Tarea AdD2	Tarea AdD13
Tarea AdD1			
Tarea AdD2			
Tarea AdD3			

# Etapa 6: Especificación de las tareas de AdD

## 1. Objetivo

- Caracterizar general de las tareas de AdD a realizar en los CA especificados en la fase anterior (objetivos, requerimientos, etc.).

## 2. Protocolo de la Fase:

- Selección y descripción de los actores y componentes necesarios para hacer cada tarea de AdD.
- Especificación de los requerimientos de c/tarea de AdD: tecnológicos, de datos, organizacionales, etc.
- Especificación de las fuentes de datos requeridas por cada tarea

# Etapa 6: Especificación de las tareas de AdD

## Tabla para describir tareas de AdD

Nombre de la tarea	<nombre de la tarea>
Descripción	<La finalidad de esta tarea>
Fuente de datos	<BD, historicos>
Tipo de tarea de analítica de datos	<Asociacion, Agrupamiento, Clasificacion, Predicción, reglas de asociación, etc.>
Técnicas de analítica de datos	<Define las posibles tecnicas a usar, por ejemplo: regresión, redes neuronales artificiales, algoritmo K-NN, etc.>
Tipo de Modelo de Conocimiento	<modelo descriptivo, modelo prescriptivo, modelo de optimizacion, modelo predictivo, etc.>
Tareas relacionadas de analítica de datos	<Con que otras tareas de AdD se relaciona>
Tipo de tarea del ciclo autonómico (rol)	<Pueden ser para observar, analizar/interpretar, o actuar sobre el proceso>

# Etapa 6: Especificación de las tareas de AdD

**Tabla para detallar especificación tareas de AdD**

Macro-Algoritmo	Especificar Tipo de Tarea de Minería	Herramienta
<paso a paso del código>	< Debe indicarse de manera concreta la tarea a realizar>	<Instrumento tecnológico a usar a utilizar para dicho calculo >
...	Por ejemplo, calcular una medida de centralidad de minería de grafo, realizar un agrupamiento de tales datos según tales criterios de similitud, etc.)	Por ejemplo, Netgraph o Netlogo para minería de grafo, o k-means para agrupamiento (indicando valor de k)
...		

**Esta tabla es particularmente importante para las tareas de AdDS**

# Fase 2: Preparación de Datos

Para aplicar AdD sobre un problema en específico, es necesario contar con un historial de datos asociado al problema en estudio.

Para realizar tareas de AdD es necesario tener los datos integrados en una sola vista, la cual comúnmente se conoce como *Vista Minable*. Existen dos tipos de vista minable:

- **Vista Minable Conceptual (VMC):** describe en detalle cada una de las variables a tomar en cuenta para c/tarea de AdD, en cada CA (proveniente de la primera fase de MIDANO).
- **Vista Minable Operativa (VMO):** Es el resultado de cargar los datos del historial y de realizar la etapa de tratamiento de datos, basado en la información de la VMC. La VMO se traduce a lo que se conoce como Vista Minable en la literatura, para realizar tareas de MD.

**Con esas vistas se construye el modelo de datos multidimensional de c/CA**

# Fase 2: Preparación de Datos

- En esta fase se plantea realizar la preparación de los datos desarrollando dos etapas. Los productos más resaltantes de esta fase son las vistas minables (conceptual y operativa) y las variables objetivos, y el modelo de datos multidimensional.



# Etapa 1: Definición del modelo de datos

## a. Objetivos

- Ubicar y comprender los datos asociados a cada tarea de AdD
- Construir una VMC que tenga las variables de interés para el caso de estudio
- Construir una VMO inicial
- Definir la(s) variable(s) objetivo(s) asociadas a los objetivos estratégicos o a responder con las tareas de AdD
- Definir el modelo de datos multidimensional de cada CA

## b. Protocolo de la etapa

- Comprender la fuente de datos de entrada
- Generar la VMC y la VMO inicial
- Integración de los datos de entrada
- Generar las tablas del modelo de datos multidimensional de cada CA

# Etapa 1: Definición del modelo de datos

## VMC

Variable	Descripción	Procedencia	Observaciones

## modelo de datos multidimensional (tipo estrella)

Nombre	Nombre de la tabla de hecho
Claves a las tablas de dimensiones	Todas las claves a las tablas de dimensiones
Variables Objetivos	Variables que describen o se asocian al conocimiento extraído (predicciones, etc.)
Otras variables	Variables requeridas por la tarea de Add, por ejemplo, derivadas de operaciones de procesamiento de las dimensiones o de OLAP

Nombre	Nombre de la tabla de dimensión
Claves de la dimensión	Clave de la dimensión
Atributos de la dimensión	Atributos que describen el tema asociado a esa dimensión

# Etapa 1: Definición del modelo de datos

## c. Productos principales

- Documento que describe las características de los repositorios donde se encuentran los datos
- Documento que describe la VMC, la cual es presentada en una tabla descriptiva.
- Vista minable operativa (modelo)
- Archivo donde esta almacenada la VMO
- Documento que describe las características de la(s) variable(s) objetivo(s )
- Modelo de datos multidimensional de cada CA
- Modelo de datos multidimensional (Constelación) del Data Warehouse

# Etapa 2: Caracterización de los datos del dominio de la aplicación

## a. Objetivos

- Construcción de la tabla con las operaciones de (E)xtracción, (T)ransformación y Carga (L), para las variables identificadas en la VMC
- Cargar los datos

## b. Protocolo de la etapa

- Integración de los datos de entrada en el DW

## c. Productos principales

- Tabla ETL

## Etapa 2: Caracterización de los datos del dominio de la aplicación

Tabla ETL

Variable	Extracción	Transformación	Carga
Nombre de la variable	De que fuente de datos organizacional se extraera	Aqjui se indican todo el proceso de pre-procesamiento de los datos (estudios de dependencia, limpieza, cambio de formatos, etc.)	A que dimensión del modelo de datos irá

**Esta tabla es de la misma longitud que la VMC**

# Etapa 3: Tratamiento de datos (ciencias de los datos)

## a. Objetivos

Esta etapa se centra en generar datos de calidad, es decir, sin anomalías, sin inconsistencias de formato, sin capturas erróneas, sin campos vacíos; aplicando métodos de limpieza, transformación y reducción sobre la vista minable operativa.

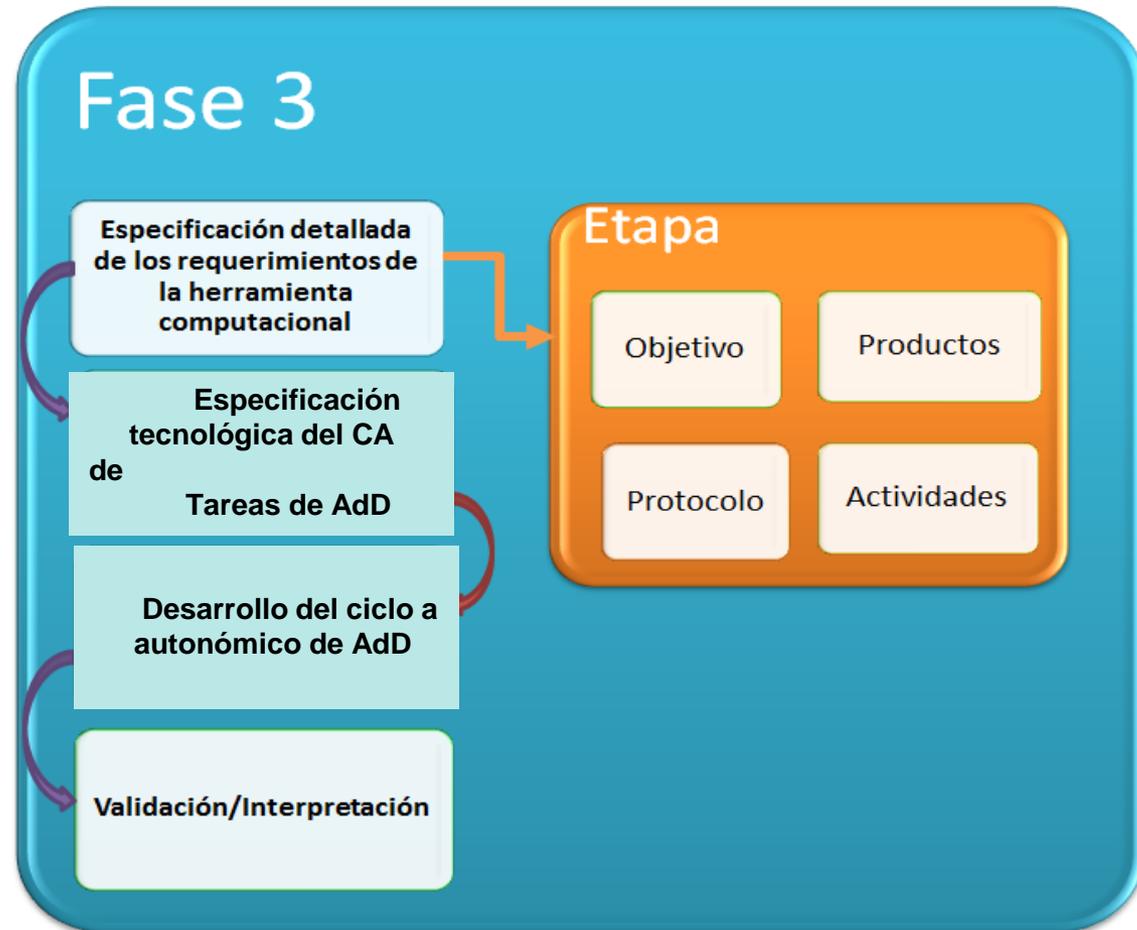
## b. Protocolo de la etapa

- Limpieza
- Transformación
- Reducción

## c. Productos principales

- VMO depurada
- DW implementada funcionalmente
- Documento descriptivo de los tratamientos realizados usando tablas descriptivas con información pertinente.

# Fase 3: Desarrollo de las tareas de AdD



# Etapa 1: Especificación detallada de los requerimientos de la herramienta computacional

## a. Objetivos

captar los requerimientos no funcionales.

## b. Protocolo de la etapa

- Requisitos de interfaz de usuario,
- Interfaces de software,
- Requerimientos de desempeño,
- Adicionalmente se pueden mencionar: de portabilidad, costos, rendimiento, accesibilidad, entre otros.

## c. Productos principales

- Informe de requerimiento no funcionales

# Etapa 2: Especificación tecnológica del ciclo autónomo de Tareas de AdD

## a. Objetivos

Caracterización la implementación tecnológica del ciclo autónomo de tareas de AdD.

## b. Protocolo de la etapa

- Escoger las técnicas de AdD para las tareas en el CA.
- Selección del Software para realizar c/tarea de AdD
- Definir cuáles son los datos de entrenamiento y de prueba contenidos en el DW a usar
- Definir las interfaces entre las tareas del CA
- Definir una estrategia para la validación de las técnicas seleccionada (cruzada, etc.).

## c. Productos principales

- Documento con la especificación tecnológica del ciclo

## Etapa 3: Desarrollo del ciclo autonómico de AdD

### a. Objetivos

Realizar la herramienta de toma de decisiones usando el ciclo autonómico de tareas de AdD.

### b. Protocolo de la etapa

- Construcción del modelo de conocimiento generado por cada tarea de AdD
- Repetir el procedimiento de ser necesario, hasta que el modelo cumpla los errores de entrenamiento establecidos
- Integrar las tareas de AdD en el CA

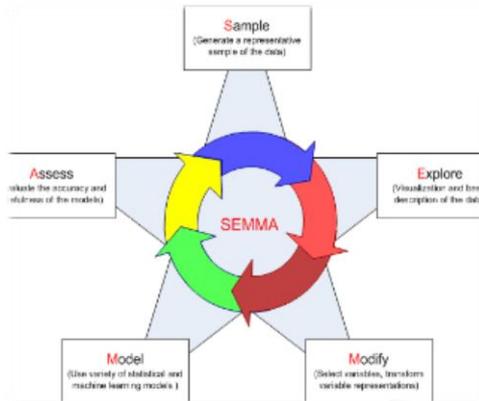
### c. Productos principales

- Prototipo del CA

**En esta etapa, se puede usar cualquier metodología de desarrollo de tareas de MD, para desarrollar las tareas de AdD.**

# Etapa 3: Desarrollo del ciclo autonómico de AdD

## Desarrollo de las tareas de AdD



### SEMMA

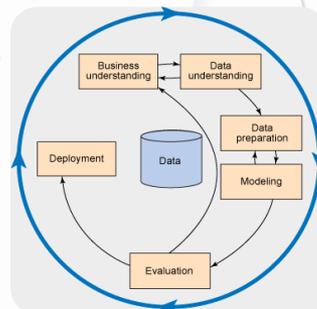
- Orientado a la parte técnica
- Carece de un análisis del problema.



Se puede usar cualquier metodología de desarrollo de MD para esta fase de desarrollo de tareas de AdD,

### CRISP-DM

- Proceso continuo y progresivo del proceso de creación
- Más utilizado por empresas que trabajan con DM



**CRISP-DM**  
CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

### CATALYST

- Estructura en “boxes”
- Primer Modelo: Analiza el problema.
- Segundo Modelo: Solución en el aspecto técnico.

# Etapa 3: Desarrollo del ciclo autonómico de AdD

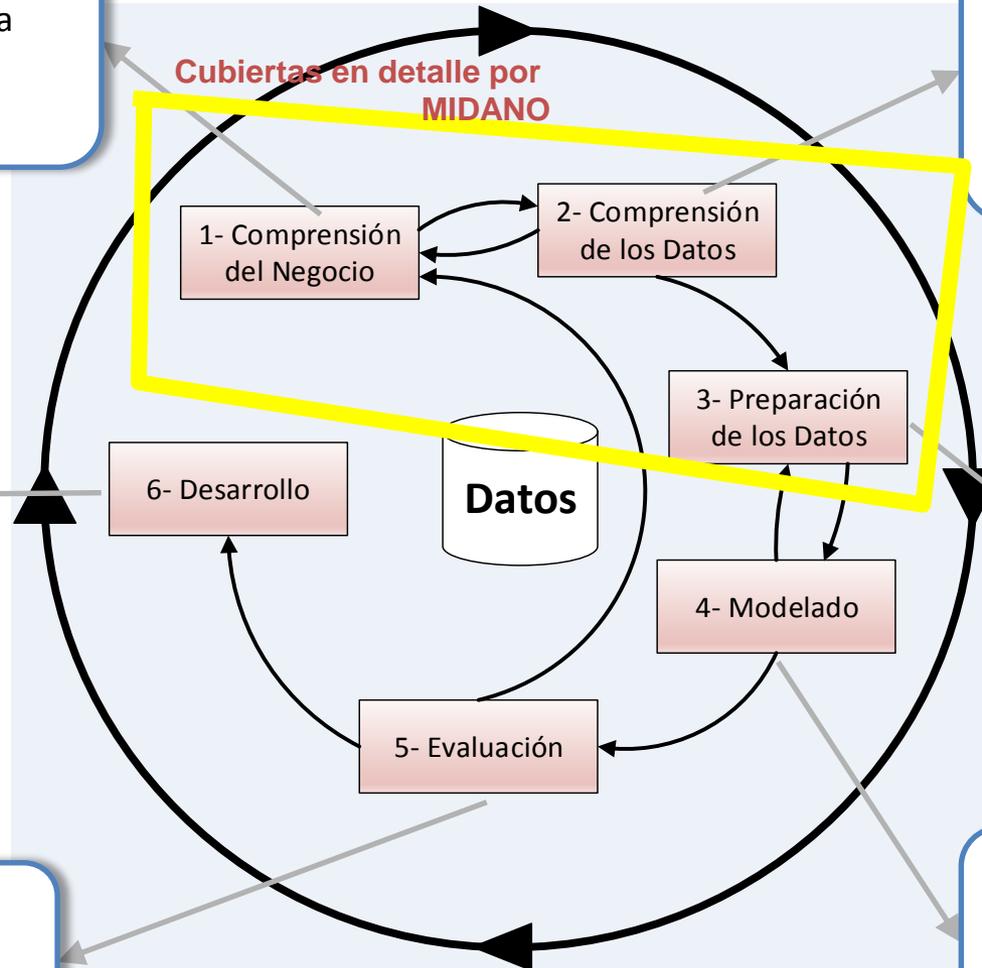
## Desarrollo de las tareas de AdD

### CRISP-DM

- Objetivos y criterios de éxito del negocio y de la MD
- Plan del Proyecto.

- Plan para el desarrollo
- Informe final
- Presentación final
- Revisión general del proyecto

- Evaluar el modelo
- Decisión sobre el modelo.



Cubiertas en detalle por  
MIDANO

1- Comprensión  
del Negocio

2- Comprensión  
de los Datos

3- Preparación  
de los Datos

4- Modelado

5- Evaluación

6- Desarrollo

Datos

- Análisis inicial de datos
- Recolección
- Descripción
- Identificación de problemas
- Verificación de calidad

- Selección de datos
- Preparar, limpiar y/o construir datos
- Generar nuevos registros
- Integrar o formatear datos

- Selección de técnica de modelado
- Obtener el modelo.

# Etapa 4: Validación/Interpretación

## a. Objetivos

Validar la herramienta de toma de decisiones.

## b. Protocolo de la etapa

- Validar el modelo de conocimiento generado por cada tarea de AdD usando los datos de prueba, y siguiendo la estrategia de validación establecida (aplicarla y observar el rendimiento).
- Realizar las correcciones necesarias
- Repetir el procedimiento de ser necesario, hasta que el modelo cumpla los errores de prueba establecidos
- Validar el comportamiento del CA, usando los criterios definidos en la etapa 1.5
- Validar el comportamiento del CA, en el sistema de toma de decisión organizacional

# yPBL, MIDANO-ext para el desarrollo de la Obra

- Metodología de base para realizar el curso: yPBL
- MIDANO-ext
  - Las tres fases permitirán especificar el problema, los datos a usar, y el diseño del ciclo atuonómico y las tareas de AdDS
- CRISP-DM
  - Las fases de desarrollo se adaptan para permitir la implementación de las tareas de AdDS

# Desarrollo del curso

## Requerimientos Funcionales

- I1: Conocer el contexto del problema
- I2: Diseño del CA de tareas de AdD
- I3: Diseño de las tareas de ADDS
- I4: Implementación y Pruebas

## Requerimientos Técnicos

- Herram. Diseño => CookBook
- Herramientas Desarrollo => cookbook
- .....=> ...



# Producto

## Esencial al curso para apropiarse del conocimiento

- Esencia del producto vs. objetivo del curso
- Contenido del curso esencial para realizar el producto
- Cuatro plantillas:

Plantillas	Uso
Definición del producto	Primera clase
Informes Técnicos	Final de cada iteración
CookBook	Final del curso
Informe Final	Final del curso (formato articulo)

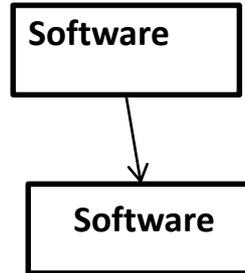
# Informes

## Definición del Producto

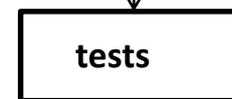
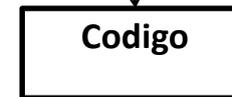
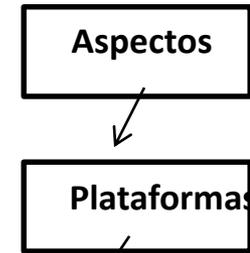
### Informes Técnicos de Avances :

- Requerimientos
- Análisis
- Diseño
- Implementación
- Pruebas

### Rama Funcional



### Rama Técnica



### Rama Desarrollo

### Cookbooks:

- Plataformas de despliegue
- Plataformas de desarrollo
- Plataforma de diseño

## Informe Final

# Plantillas

## Definición del Producto

- **Nombre**
- **Objetivo**
- **Descripción del problema**
- **Interés del problema**
- **Palabras claves**
- **Posibles Clientes/Dolientes**

# Plantillas

- **Informes Técnicos de Avances**
  - Objetivo de la iteración
  - Caracterización en el producto
  - Diseño en el producto
  - Prototipo y pruebas

# Plantillas

## CookBook (>10% nota final)

- **Resumen (Abstract)**
- **Palabras Claves (Keywords)**
- **Contribuyentes (Contributors)**
- **Versiones (Releases)**
- **Introducción (Introduction)**
- **Ingredientes: Definiciones y Terminología (The ingredients: Definitions and terminology)**
  - **Ingrediente 1 (Ingredient 1)**
- **Recetas (Recipes)**
  - **Receta 1: Una primera receta (Recipe1: A first recipe (e.g. a HelloWorld recipe))**
    - **Paso 1: descripción paso 1 (Step1: short description of step 1)**
- **Documentación Recomendada (Recommended documentation)**
- **Referencia 1 (Reference 1)**
- **Retroalimentación (Feedback)**

# Plantillas

## Informe Final

- **Introducción**
- **Marco Teórico específico para el Producto**
- **Presentación del producto**
- **Diseño del producto**
- **Desarrollo del producto**
- **Prototipos y pruebas**
- **Conclusiones**

**Formato IEEE**

# Posibles Proyectos

- **Nombre**
- **Objetivo**
- **Descripción del problema**
- **Interés del problema**
- **Palabras claves**
- **Relación con el curso**
- **Expectativas con el curso**

# Posibles Proyectos

**Analizar y procesar grandes volúmenes de datos en el contexto de interacciones estudiantiles**

especialmente en entornos de cursos abiertos online, por ejemplo MOOCs

**Análisis de grandes volúmenes de datos a través de enfoques basados en Web Semántica**

analizar contenido no estructurado sobre corpus basados en microblogging.

**Big Data y Minería de datos**

Integración de los niveles de conocimiento con las herramientas BigData

**Descripción y Representación de Cuerpos de Conocimiento (BOK)**

Determinar y establecer los datos que permitan una correcta descripción de un BOK

**Validar si los recursos en el EVA ayudan en el aprendizaje en matemática**

Los docentes envían/cargan recursos educativos que no son utilizados por el estudiante pero no son útiles. Conocer técnicas para ayudar con este tipo de problemas

**Medir la influencia del profesor en el rendimiento académico**

de los estudiantes de la asignatura de estadística de la UTP

**Análisis del rendimiento académico de los estudiantes** de la asignatura lógica de la programación de la modalidad a distancia

**actualizar el modelo ontológico con fuentes externas, tesauros,**

los resultados de la fase de comparación que se encuentran en una base de datos, la generación de conocimiento desde el artículo 1 y el análisis de habilidad.

# Posibles Proyectos

## Web semántica y Linked Data

La falta de interconexión entre diferentes conjuntos de datos de dominios similares

## Aprender a extraer información de IoT

Hacer más interactivos los objetos que utilizamos en nuestra vida diaria

**Aprendizaje colaborativo aplicable a sistema emergente** de repositorio inteligente de recursos didácticos validar recursos para el aprendizaje colaborativo

**Analizar cómo influye la recomendación** de recursos en el aprendizaje de los estudiantes

**Análisis de la influencia de actividades síncronas y asíncronas** en el rendimiento académico en el rendimiento académico de los estudiantes

**Identificar consenso entre los datos sociales** presentes en SaCI para mejorar el proceso de enseñanza-aprendizaje

**Generar indicadores a partir de los datos del Entorno Virtual de Aprendizaje**

Falta de indicadores que ayuden a la toma de decisión en cuanto a cómo los estudiantes aprenden

**Identificar tendencias de interrogantes generadas en el EVA** en el proyecto de Asistente Virtual para respuestas automáticas

# Ejemplo de uso de Midano

# Fase 1: Conocimiento de la Organización

## Caso de Estudio: Empresa Petrolera

### *Etapa 1: Conocimiento de la organización:*

Se trata de una empresa que se encarga de la exploración, extracción, producción, mejoramiento y comercialización de crudo extrapesado.

### *Etapa 2: Caracterización de los procesos de la organización*

La cadena de valor de la empresa se muestra en la siguiente figura, donde el proceso principal objeto de estudio se concentra en la tercera etapa de la cadena de valor.



Para el grupo de expertos, una de las etapas más importantes para obtener el producto deseado es la refinación, llevada a cabo en lo que se conoce como “complejo mejorador”.

# Fase 1: Conocimiento de la Organización

## Caso de Estudio: Empresa Petrolera

### *Etapa 3: Selección del Proceso*

Se estudió cada uno de los subproceso (objetivos, actividades, productos, etc.), y se obtuvo la interacción entre ellos.

En la tabla se ilustra este proceso de priorización y selección, considerando sólo los dos procesos que resultaron mejor ponderados en este caso de estudio.

Crterios	CDU	DCU
Importancia para la organización	5	5
Propósito de la MD	5	5
Interacciones entre procesos	2	4
Procesos dependientes	5	3
Importancia de la calidad del producto	4	4
Seguridad Industrial	4	5
Replicabilidad de la herramienta desarrollada	5	4
Cantidad de Expertos	5	5
Fuentes de información	5	5
Confidencialidad de la información	3	3
¿Qué información se recoge del proceso para ser almacenada?	5	5
Con que frecuencia se recoge la información almacenada	4	4
¿Qué herramientas se cuentan, para recolectar y manipular la información?	4	4
Total sin ponderación	56	56
Total ponderado	83	76

### *Descripción del escenario futuro*

El escenario futuro seleccionado es para **predecir la calidad del producto y optimizar la cantidad de nafta a la salida de la columna destilador atmosférico.**

# Fase 1: Conocimiento de la Organización

## Caso de Estudio: Empresa Petrolera

### *Etapa 4: Análisis para caracterizar las posibles tareas de Minería de Datos (MD)*

#### Descripción del escenario actual

Resultados que se obtienen	Actor(es) asociado(s)	Variables Asociadas	Actividades que se realizan
<b>Gasoil directo (SRGO), nafta pesada y residuo atmosférica.</b>	<ul style="list-style-type: none"><li>• Expertos asociados al proceso</li><li>• Ingenieros de Procesos</li><li>• Operadores</li><li>• Unidad de destilación atmosférica</li></ul>	<ul style="list-style-type: none"><li>• Tren de precalentamiento: temperatura de la carga.</li><li>• Desaladores: tiempo para el asentamiento y separación del agua del petróleo, presión.</li><li>• Hornos de crudo: temperatura</li><li>• Columna de crudo: presión, temperatura, rata de vapor de despojamiento.</li></ul>	<ul style="list-style-type: none"><li>• Carga del crudo.</li><li>• Precalentamiento del crudo diluido.</li><li>• Desalado.</li><li>• Precalentamiento del crudo desalado.</li><li>• Generación de cortes de crudo en la columna.</li></ul>

# Fase 1: Conocimiento de la Organización

## Caso de Estudio: Empresa Petrolera

### Descripción del escenario futuro

Resultados que se desean obtener	Actor(es) asociado(s)	Variables Asociadas	Actividades de MD que se realizarían	Funcionalidades nuevas
<b>Predicción de la calidad del producto, para optimizar el proceso</b>	<ul style="list-style-type: none"><li>• Expertos asociados al proceso</li><li>• Operadores</li><li>• Columna de crudo</li></ul>	Presión, temperatura de tope y rata de vapor de despojamiento de la columna de crudo.	Predicción	<ul style="list-style-type: none"><li>• Predicción de las características del producto, según las condiciones de funcionamiento de la torre de crudo.</li><li>• Ayudar a optimizar el proceso de producción, generando información para orientar a los actores en la toma de decisiones con la predicción (es) resultante(s).</li></ul>

### *Descripción del escenario futuro*

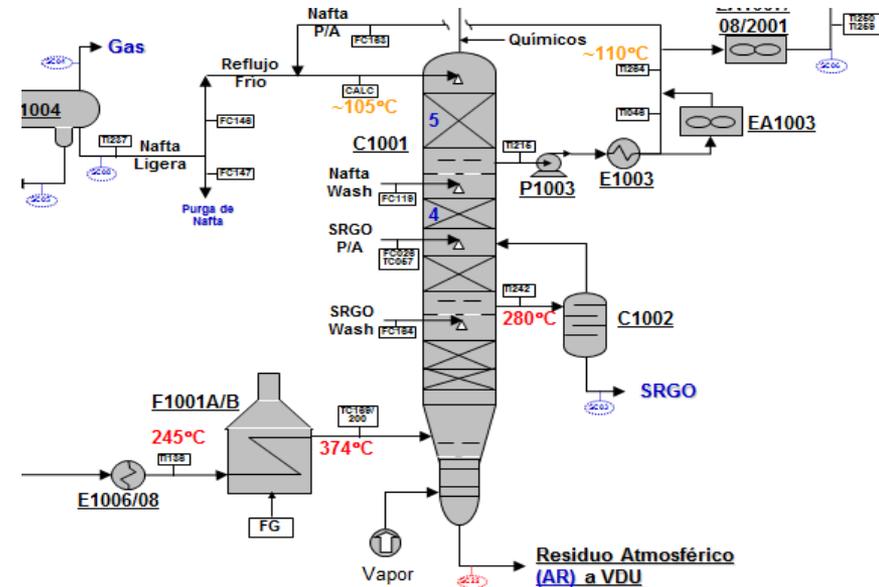
El escenario futuro seleccionado es para **predecir la calidad del producto y optimizar la cantidad de nafta a la salida de la columna destilador atmosférico.**

# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- **Comprensión de los datos de entrada**

A través de los diagramas de instrumentos de la planta, se determinaron cuáles son los datos asociados a las variables, se tomaron las variables más importantes asociadas al escenario futuro



# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- **Construcción de la VMC**

A partir del escenario futuro escogido, y con apoyo del grupo de expertos, se construyó una VMC con las características de la Tabla 1. Debido a que la misma cuenta con más de cien variables, sólo se presenta una pequeña muestra.

Tabla 1. Muestra de la VMC del escenario seleccionado

Variable	Descripción	Dependencia	Observaciones
11FC900	Flujo de nafta de lavado para la preflash	Identificar relación con FY119	Controlada
11PI1005	Presión de entrada de nafta wash	-	No es relevante para el estudio
11PI001A	Presión tope de la columna preflash	-	-

# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- ***Construcción de la VMO***

Se carga el historial en un archivo con las variables obtenidas en la VMC.

Los datos proporcionados por la empresa fueron entregados en formato Excel, donde todos los datos están integrados en un documento menos una variable de laboratorio, ya que la misma, es tomada con una frecuencia diferente a las demás variables.

# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- *Integración de los datos de entrada*

Fecha	11_FI158T_PNT	....	11_FC010_MEAS
01.01.2009 0:00:00	320.139504	....	39.03201294
01.01.2009 0:05:00	318.8554796		39.03201294
01.01.2009 0:10:00	315.9257853		39.03201294
01.01.2009 0:15:00	316.9394877		39.03201294
01.01.2009 0:20:00	316.2324899		39.03201294
01.01.2009 0:25:00	318.2673392		39.03201294
01.01.2009 0:30:00	311.0020414		39.03201294
01.01.2009 0:35:00	314.7039024		39.03201294
.	.	.	.
.	.	.	.

(a) Formato de la tabla de datos con las variables asociadas a los sensores de la planta

Fecha	[°API]
02.07.2008 05:00:00	45.9
03.07.2008 05:00:00	46.1
04.07.2008 05:00:00	46.1
05.07.2008 05:00:00	46.2
06.07.2008 05:00:00	46.4
07.07.2008 05:00:00	45.8
08.07.2008 05:00:00	46
09.07.2008 05:00:00	45.6
10.07.2008 05:00:00	45.1
11.07.2008 05:00:00	45.4
12.07.2008 05:00:00	45.3
13.07.2008 05:00:00	45.6
.	.
.	.

(b) Formato de la tabla de datos de gravedad API (medición de laboratorio)

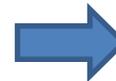
# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- **Construcción de la VMO**

	11_FI158T_PNT	51_FT006_PNT
02.01.2009 23:30:00	314.7672055	1393.200177
02.01.2009 23:35:00	313.6730738	1396.853361
02.01.2009 23:40:00	317.3760808	1391.633283
02.01.2009 23:45:00	314.5253747	1391.253645
02.01.2009 23:50:00	315.1430386	1398.356516
02.01.2009 23:55:00	311.9205457	1400.912088
03.01.2009 0:00:00	312.5063793	1392.555884
03.01.2009 0:05:00	312.7566352	1394.478128
03.01.2009 0:10:00	312.8069345	1399.825388
03.01.2009 0:15:00	312.0659453	1401.837267

	Gravedad API a 60 °F [*API]
02.01.2009 05:00:00	47
03.01.2009 05:00:00	46.7
04.01.2009 05:00:00	46.8
05.01.2009 05:00:00	47.1
06.01.2009 05:00:00	48.6
07.01.2009 05:00:00	46.9



## Vista minable operativa(VMO)

	11_FI158T_PNT	51_FT006_PNT	[*API]
02.01.2009 23:30:00	314.7672055	1393.200177	47
02.01.2009 23:35:00	313.6730738	1396.853361	47
02.01.2009 23:40:00	317.3760808	1391.633283	47
02.01.2009 23:45:00	314.5253747	1391.253645	47
02.01.2009 23:50:00	315.1430386	1398.356516	47
02.01.2009 23:55:00	311.9205457	1400.912088	47
03.01.2009 0:00:00	312.5063793	1392.555884	46.7
03.01.2009 0:05:00	312.7566352	1394.478128	46.7
03.01.2009 0:10:00	312.8069345	1399.825388	46.7
03.01.2009 0:15:00	312.0659453	1401.837267	46.7

# Fase 2: Preparación de datos

## Etapa 1. Dominio de la aplicación

- **Definir las variables objetivo**

Observar el(los) objetivo(s) de cada una de las variables en el escenario futuro seleccionado. Con el escenario futuro seleccionado se obtiene lo siguiente:

- Escenario futuro: Producir la mayor cantidad de Nafta a 46 API
- Funcionalidades nuevas: Predicción del API del producto, según las condiciones de funcionamiento de la torre de destilación.

Tabla 2. Variables objetivos

Variables objetivo	Observaciones
API NAFTA	Predecir el api de la nafta
11FC158	Maximizar el flujo de nafta producto a 46 api

# Fase 2: Preparación de datos

## Etapa 2. Tratamiento de Datos

- Limpieza**

Se ubicaron las variables con más errores en la VMO. Los resultados obtenidos son reflejados en la Tabla 3.

Tabla 3. Variables con mas errores en la VMO

PERIODO	Ene-Mar 2008	Abril-Jul 2008	Jul-Sept 2009	Oct-Dic 2010	Abril-Jul 2011	Ener- Marzo 2012	Oct-Dic 2013
NOMBRE DE LA VARIABLE	% Error	% Error	% Error	% Error	% Error	% Error	% Error
11_FC158_MEA S	0,00635 93	92,4336 1416	100	100	100	100	100
11_FC044_MEA S	15,4276 6296	60,7603 7112	22,4308 8159	100	100	0,00485 2014	100
11_FC300_MEA S	0,00635 93	18,7693 2921	4,47872 4197	100	100	100	100
11_FC119_MEA S	0,00635 93	39,4315 8793	4,47872 4197	100	100	100	100
11_FC133_MEA S	99,9936 407	99,7493 868	100	0,00546 38	14,6290 1772	33,9155 7496	0,00771 5454

# Fase 2: Preparación de datos

## Etapa 2. Tratamiento de Datos

- **Limpieza**

La Tabla fue evaluada con los expertos del proceso con la finalidad de definir acciones que se podrían tomar. Las acciones tomadas son resumidas en la tabla 4, donde se describe la justificación de cada acción realizada.

Tabla 4. Acciones tomadas con las variables con más anomalías en la VMO

NOMBRE	JUSTIFICACIÓN	ACCIÓN
11_FC158_MEAS	Se puede eliminar porque es el mismo registro que el 11_FI158T_PNT	Eliminar de la VMO
11_FC044_MEAS	Se puede eliminar del estudio (es una línea de arranque o usada en operaciones muy puntuales).	Eliminar de la VMO
11_FC300_MEAS	Preferiblemente incluirla, si esta 11_FT300_PNT, se puede eliminar	Estudiar
11_FC119_MEAS	Preferiblemente incluirla, si esta FC119, se puede eliminar	Estudiar
11_FC133_MEAS	Se puede eliminar	Eliminar de la VMO

# Fase 2: Preparación de datos

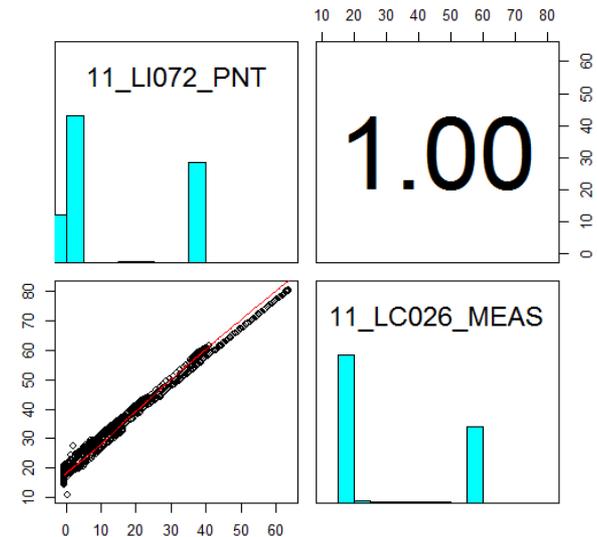
## Etapa 2. Tratamiento de Datos

- **Limpieza**

Para las variables que tienen dependencias con otras variables, se construyeron modelos lineales para sustituir los datos dañados por el resultado de estas relaciones.

Dichas dependencias pueden ser apoyadas y justificadas usando un gráfico de dispersión. En este caso, se seleccionó un diagrama de dispersión con las siguientes características:

- El histograma y nombre de cada variable (ver la diagonal de la Figura).
- La distribución de los puntos entre las dos variables y la curva regresada (parte inferior izquierda de la Figura).
- El coeficiente de correlación entre parejas de variables (parte superior derecha de la Figura).



- **Transformación**

En este estudio no aplica el proceso de transformación, debido a que toda la data se encuentra en un formato consistente de unidades y magnitudes en las variables.

# Fase 2: Preparación de datos

## Etapa 2. Tratamiento de Datos

- Reducción**

Se realizaron análisis estadísticos entre variables que el experto identificó con dependencias en la VMC y además se construyó una matriz con la correlación entre todas las variables, con la finalidad de identificar las variables altamente correlacionadas.

	11_FI158T_PNT	11_FIC011_PNT	11_FC010_MEAS	11_FC159_MEAS	11SC01_BSW_NUM	11SC24_BSW_NUM
11_FI158T_PNT	1	0.368111972669728	0.484318010844852	-0.204051244869186	0.144851822147737	-0.311201934354127
11_FIC011_PNT	0.368111972669728	1	0.210678189668377	0.180986820619784	0.270251411195036	-0.067983791041037
11_FC010_MEAS	0.484318010844852	0.210678189668377	1	0.0075735692671944	-0.066805623663014	0.0505010255310379
11_FC159_MEAS	-0.204051244869186	0.180986820619784	0.0075735692671944	1	0.373898847616421	0.865146726620298
11SC01_BSW_NUM	0.144851822147737	0.270251411195036	-0.066805623663014	0.373898847616421	1	0.340144245505393
11SC24_BSW_NUM	-0.311201934354127	-0.067983791041037	0.0505010255310379	0.865146726620298	0.340144245505393	1
11_PI1001A_PNT	0.149797893568095	-0.230486068559374	0.0795532368827437	-0.530813442822542	-0.169561318779577	-0.265363516836238
11_PDI1001_PNT	-0.3513783530082	-0.064687446799048	0.0146327207575092	0.898505092755542	0.280231203226639	0.987591989101072
11_FC069_MEAS	0.836782714193593	0.429652170296929	0.412503957853951	-0.246747320588408	0.069211016952507	-0.427705464814407
11_TC167_MEAS	0.540859492533276	0.254940613271293	0.297113202094931	0.136683941679193	0.121901253006149	0.0642377397275748
11_TI168_PNT	0.528217711006828	0.251589322326238	0.294849147655103	0.16455588536907	0.131377400543795	0.0938984439071776
11_FC097_MEAS	0.863314754277501	0.4603093266029	0.453315700501899	-0.146716637712636	0.198073044334871	-0.327059949655631
11_TI205_PNT	0.465570226005517	0.231037998971888	0.128340774528971	-0.247838628436702	-0.262645490028422	-0.465553704306329
11_PI149_PNT	-0.023605088600139	0.24201692083265	0.0128921387555581	0.479588931261072	-0.030646242732858	0.183565008277804
11_PDI148_PNT	0.290003045342975	-0.090345184479632	0.0422040129610041	-0.978816278934173	-0.367572420350681	-0.9111512964571754
11_FC164_MEAS	0.520628675400118	0.241198936667788	0.327886765160563	-0.021969698452506	-0.060401548555582	-0.074430658678374
11_FC114_MEAS	0.669735791599634	0.377566765068302	0.232778574068234	-0.481099372744771	0.152729477350047	-0.612534965210986
11_TI206_PNT	0.379966415482969	0.0902448084933138	0.0880606872816109	0.68915732489563	-0.431286570684153	-0.781755090555928
11_LC016_MEAS	0.210108572264961	0.265240077290013	0.125599091378307	0.366193900688599	0.410047378697821	0.228435524378336
11_FC148_MEAS	-0.55524114311874	-0.255163977546328	-0.316329956267743	-0.083064170826386	-0.187975808467312	-0.041865764739969
11_FI149_PNT	-0.429285166720203	-0.103822072026452	-0.107453649503538	0.80189622301047	0.150822460062886	0.833467964006083
11_TI204_PNT	0.716377153818456	0.303070770590223	0.344980429883973	-0.184537891817327	0.009581408992425	-0.2871877823792

Se realizó la reducción de la VMO sobre las variables con alta correlación. La VMO inicial contaba con 97 variables, después de realizar la limpieza y reducción de datos la VMO final cuenta con 33 variables.

# Fase 2: Preparación de datos

## Etapa 2. Tratamiento de Datos

- **Reducción**

Tabla 5. Reducción de Variables

Nombre	Resultado	Acción	Justificación
11_PI1001B_PNT	Correlación alta con 11_PI1001A_PNT	ELIMINAR	Variable altamente correlacionada, donde ambas aportan la misma información
11_PI157A_PNT/ 11_PI157B_PNT	Correlación alta con 11_PI149_PNT	ELIMINAR	Variable altamente correlacionada, donde las tres aportan la misma información
11_PI156_PNT	Correlación alta con 11_PI155_PNT	ELIMINAR	Variable altamente correlacionada, donde ambas aportan la misma información
· · ·			

Se realizó la reducción de la VMO sobre las variables con alta correlación. La VMO inicial contaba con 97 variables, después de realizar la limpieza y reducción de datos la VMO final cuenta con 33 variables.

# Fase 3: Desarrollo de herramientas de MD

## Etapa 1: Especificación detallada de los requerimientos de la herramienta computacional

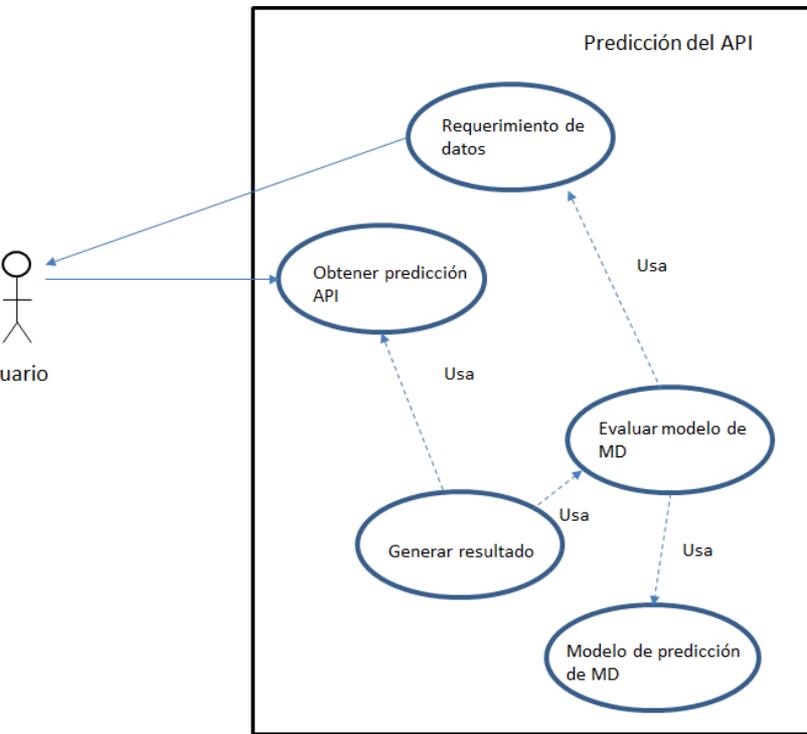
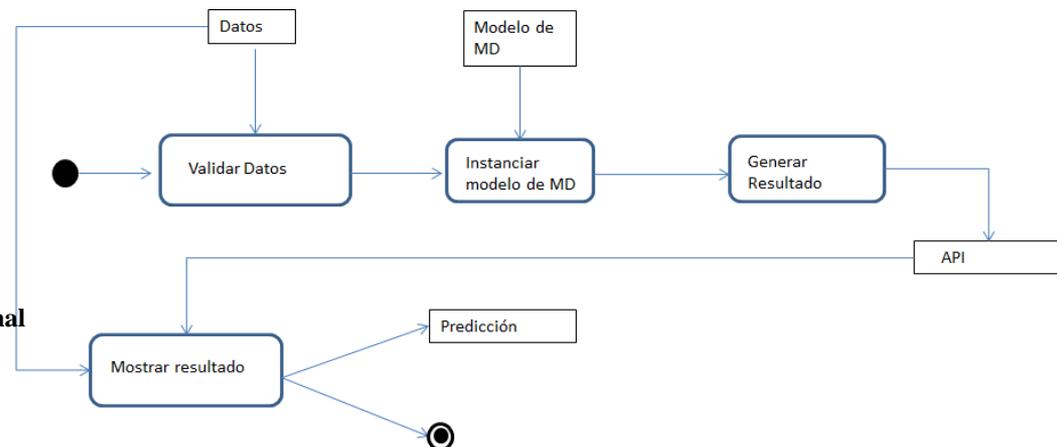


Diagrama de actividades de la herramienta computacional para la predicción del API

### Entre los requisitos no funcionales:

- Requisitos de interfaz de usuario.
- Interfaces de software.
- Requerimientos de desempeño.
- Otros: de portabilidad, costos, accesibilidad, etc.

### Caso de uso para la predicción del caso de estudio



# Fase 3: Desarrollo de herramientas de MD

## Etapa 2: Desarrollar el modelo de MD

### Predicción del API de nafta:

1. Selección del Software para realizar las tareas de MD: Weka.
2. Escoger la tarea de MD para el escenario futuro: predicción.
3. Definir cuáles son los datos de entrenamiento y de prueba dispuestos en la vista minable: Los datos de entrenamiento son el 70% de los registros de la VMO y el resto será utilizado para realizar una validación cruzada<sup>2</sup> con Weka
4. Selección del algoritmo de MD: Los algoritmos considerados a evaluar son en general algoritmos basados en regresión lineal y redes neuronales.

Tabla 6. Algoritmos evaluados para la predicción del API de nafta

Algoritmo	Mean absolute error	Root squared error
LinearRegression	0.3546	0.503
RBFNetwork	0.4964	0.6953
SimpleLinearRegression	0.4422	0.6198
PaceRegression	0.3562	0.5028
IsotonicRegression	0.4293	0.5923

7. Modelo de MD: modelo que expresa las relaciones, a través de fórmulas y reglas, entre las variables del proceso. La ecuación obtenida para la predicción del API obtenida con *LinearRegression* viene dada por

$$\text{API} = 0.0032 * 11\_FI158T\_PNT + 0.0006 * 11\_FIC011\_PNT - 0.004 * 11\_FC010\_MEAS - 0.1673 * 11SC01\_BSW\_NUM + 2.8329 * 11\_PI149\_PNT - 0.1857 * 11\_FC114\_MEAS + 0.0481 * 11\_TI206\_PNT - 0.0027 * 11\_LC016\_MEAS - 0.0087 * 11\_FC148\_MEAS + 0.0042 * \dots$$

Validar



# Fase 3: Desarrollo de herramientas de MD

## Etapa 3: Implementación usando el modelo de MD

### ○ Formato de la entrada

	A	B	C	D	E	F	G	H	I	J
1	Fecha	11_FI158T_P	11_FIC011_P	11_FC010_M	11SC01_BSW	11SC24_BSW	11_TC167_M	11_TI205_PN	11_PI149_PN	11_FC114_M
2	-	303.373624	156.608638	40.409008	2.23353173	0.03146036	376.358974	370.675293	1.15274024	7.02434635
3	-	303.980151	151.507827	40.409008	2.23392856	0.03146106	376.532866	370.675293	1.15274024	7.02434635
4	-	304.417303	142.997793	40.409008	2.23432538	0.03146176	376.694006	370.675293	1.15274024	7.02434635
5	-	297.576434	141.570275	40.409008	2.23472221	0.03146246	376.845872	370.675293	1.15274024	7.02434635
6	-	301.650863	142.512228	40.409008	2.23511903	0.03146316	377.019765	370.675293	1.15274024	7.02434635
7	-	300.597413	161.583226	40.409008	2.23551586	0.03146386	377.193657	370.675293	1.15274024	7.02434635
8	-	300.120383	240.557088	40.409008	2.23591268	0.03146456	377.367549	370.675293	1.15274024	7.02434635
9	-	295.659939	244.557774	40.409008	2.23630951	0.03146526	377.541442	370.675293	1.15274024	7.02434635
10	-	296.365927	248.684239	40.409008	2.23670634	0.03146596	377.690216	370.675293	1.15274024	7.02434635
11	-	299.257376	243.036145	40.409008	2.23710316	0.03146666	377.587813	370.675293	1.15274024	7.02434635
12	-	297.409396	244.189429	40.409008	2.23749999	0.03146736	377.413921	370.675293	1.15274024	7.02434635
13	-	299.060158	243.92957	40.409008	2.23789681	0.03146806	377.240028	370.675293	1.15274024	7.02434635

Variables de la VMO

Prototipo para predecir el API de nafta

Abrir

Aceptar Nuevo

Evaluar Salir

### ○ Salida

Prototipo para predecir el API de nafta

C:\Users\Fannia\Desktop\prueba.xls

Abrir

Aceptar Nuevo

VM Prediccion

Fecha	API
-	47.07878...
-	47.08143...
-	47.07756...
-	46.84269...
-	46.84269...

Evaluar Salir

# Ejemplo de uso de MiDANO-ext en SaCI

SaCI

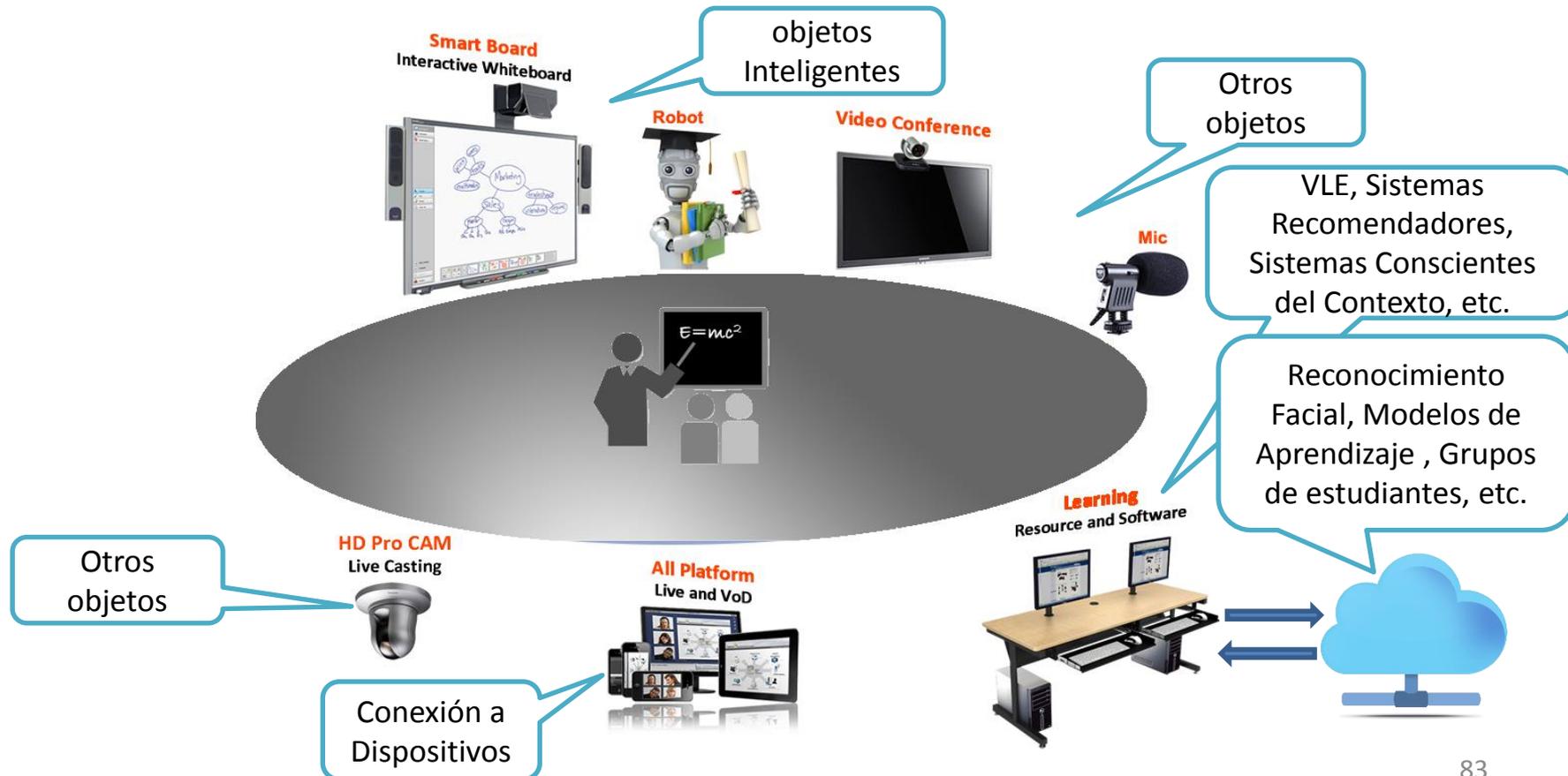
# Aula inteligente

Espacio donde la tecnología  
ubicua ayuda al **proceso de  
enseñanza-aprendizaje** de una  
manera transparente.



# Especificación de SaCI

## Componentes de un aula inteligente





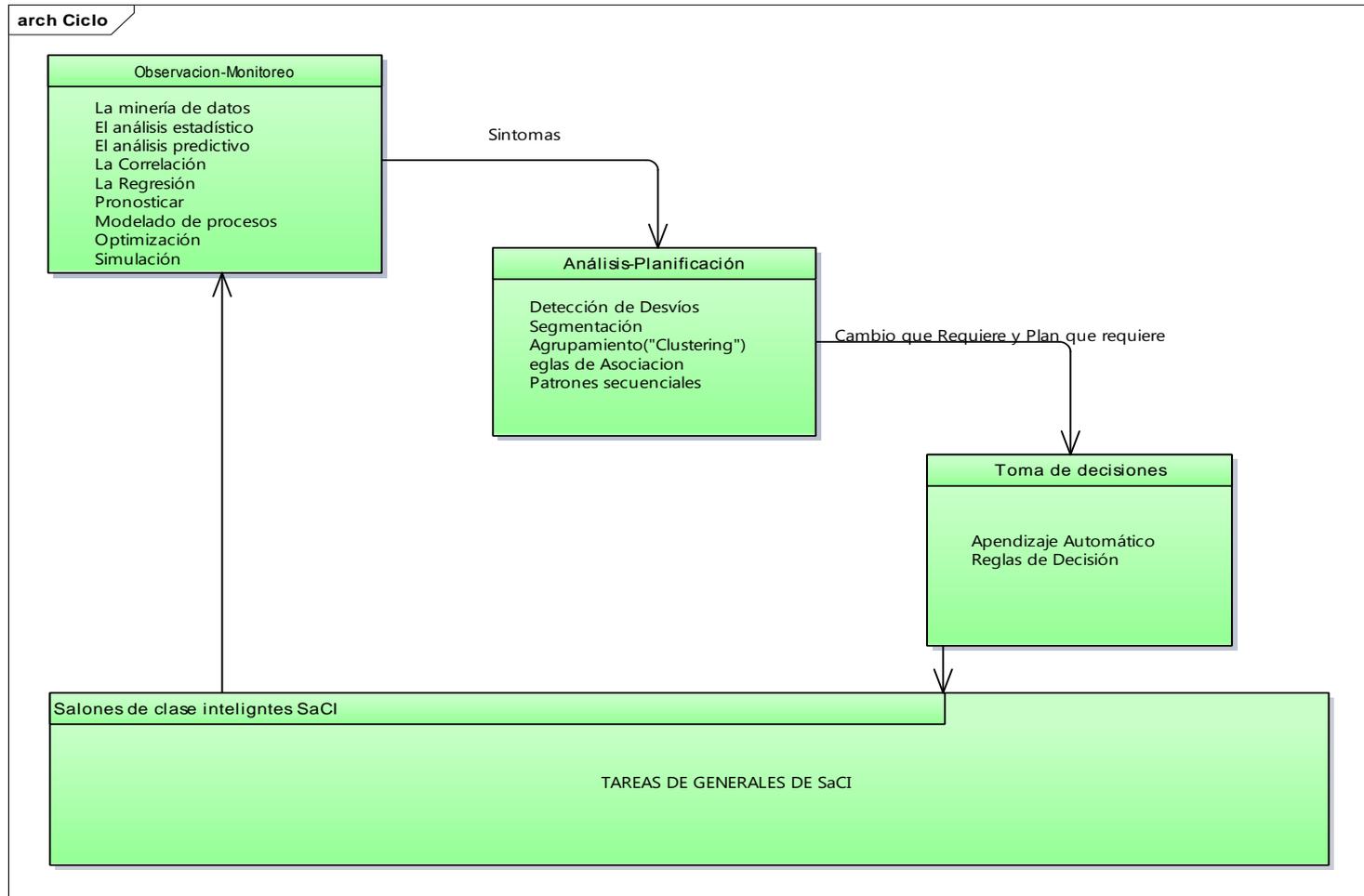
## Objetivo AdD

Entender y mejorar el proceso de enseñanza y aprendizaje.

## Objetivos específicos

- Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).
- Identificar los recursos educativos ideales para un estudiante en un momento dado.
- Identificar los estudiantes que necesitan más atención y sus necesidades.
- Evitar la deserción de estudiantil.

# Especificación del Ciclo Autónomo



# Especificación del Ciclo Autónomo

## Ciclo 1

**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

## Ciclo 2

**Objetivo:** Identificar los recursos educativos ideales para un estudiante en un momento dado.

## Ciclo 3

**Objetivo:** Identificar los estudiantes que necesitan más atención y sus necesidades.

## Ciclo 4

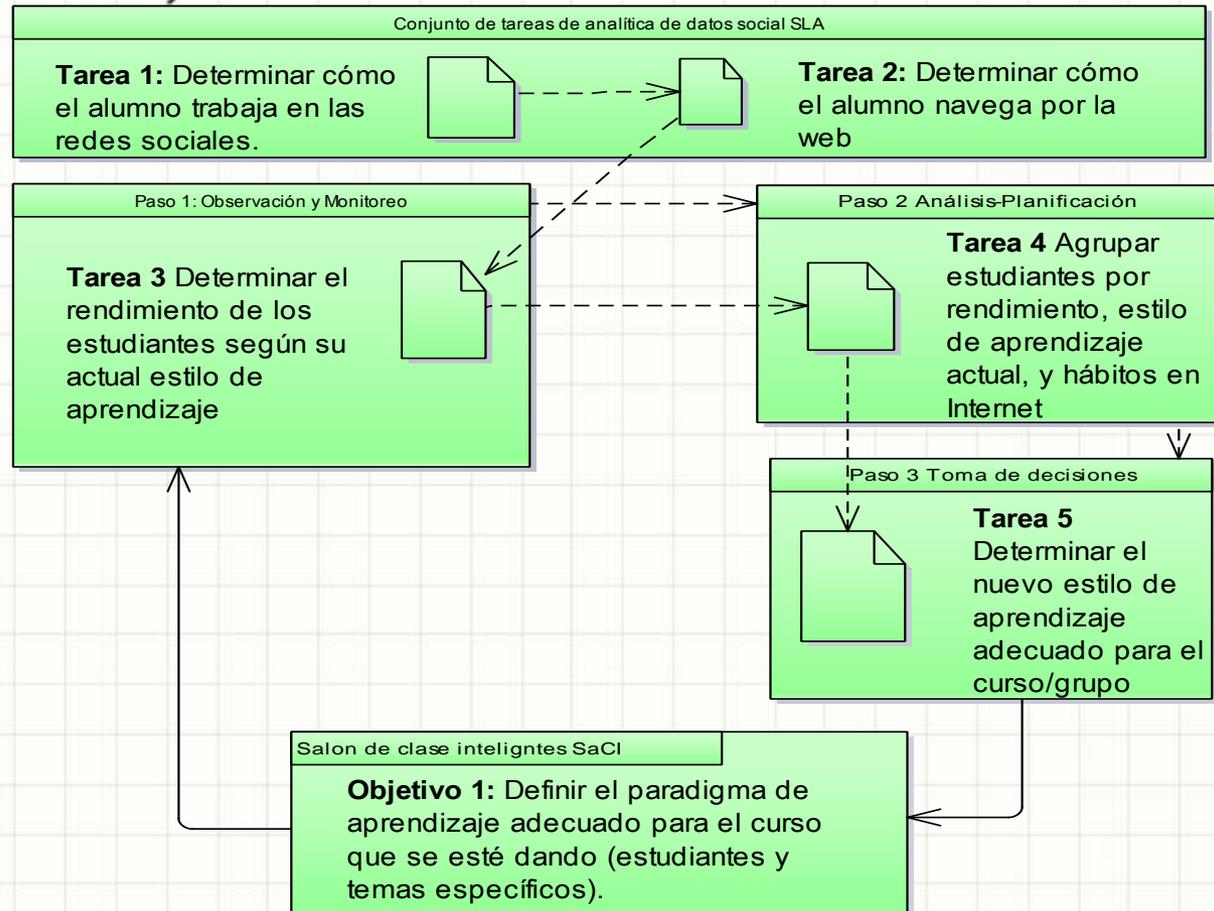
**Objetivo:** Evitar la deserción de estudiantil.

# IMPLEMENTACIÓN DEL CICLO AUTONÓMICO DE TAREAS.

## Ciclo 1

dfd Ciclo\_objetivo\_1

**Ciclo 1: Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).**



**MODELO  
DEL  
CICLO 1**

# Implementación del Ciclo autónomo de Tareas.

## Ciclo 1

**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

**Paso 1: (Observación)**

**Tarea 1/5:** Determinar cómo el alumno trabaja en las redes sociales.

SaCI

Resultados Esperados	Actores	Variables	Herramienta Social	Actividades
Obtener el grafo del comportamiento del estudiante en redes sociales	Estudiantes	Estudiantes Redes sociales	Minería de grafos	Obtener el grafo de comportamiento
<b>Nombre de la Tarea</b>		<b>Determinar cómo el alumno trabaja en las redes sociales</b>		
<b>Descripcion</b>		Determinar el grafo e comportamiento del estudiante en las redes sociales.		
<b>Fuente de Datos</b>		Redes sociales		
<b>Tipo de tarea de analítica de datos social</b>		Minería de grafos		
<b>Técnica de analítica social</b>		Cascada Independiente / Influence Maximization		
<b>Tarea con la que se relaciona</b>		 <pre> graph LR     A[Minería de grafos.] --&gt; B[Minería de uso]             </pre>		
<b>Tipo de tarea en el ciclo</b>		Observacion		

# Implementación del Ciclo autonómico de Tareas.

## Ciclo 1

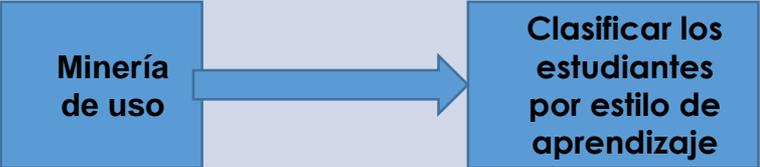
**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

**Paso 1: (Observación)**

**Tarea 2/5:** Determinar cómo el alumno navega por la web

Resultados Esperados	Actores	Variables	Herramienta Social	Actividades
Obtener un patrón de navegación del estudiante en internet	Estudiantes	Estudiantes Páginas web	Minería de uso	Obtener el patrón de Navegación en internet

SACI

Nombre de la Tarea	Minería de uso para determinar como el alumno navega por la web
Descripción	Determinar la las preferencias del estudiante con relación a la web por medio de las páginas que visita y así obtener el estilo.
Fuente de Datos	Internet
Tipo de tarea de analítica de datos social	Minería de uso.
Técnica de analítica social	PageRank
Tarea con la que se relaciona	 <pre> graph LR     A[Minería de uso] --&gt; B[Clasificar los estudiantes por estilo de aprendizaje]             </pre>
Tipo de tarea en el ciclo	Observación

Ciclo autonómico de tareas para

# Implementación del Ciclo autonómico de Tareas.

## Ciclo 1

**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

**Paso 1: (Observación)**

**Tarea 3/5:** Determinar el rendimiento de los estudiantes según su actual estilo de aprendizaje

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Determinar el rendimiento de los estudiantes por estilo de aprendizaje utilizado.	Estudiante	Estilo de A. Estudiante	Clasificar	Clasificar estudiantes de un Estilo de Aprendizaje por rendimiento

Nombre de tarea	Determinar el rendimiento de los estudiantes según su actual estilo de aprendizaje
Descripción	Identificar los tipos de rendimientos de los estudiantes de acuerdo al estilo de aprendizaje utilizado
Fuente de datos	Base de Datos de SaCI
Tipo de tarea de analítica de datos	Clasificación
Técnica de analítica de datos	
Con que otras tareas se Relaciona	<pre> graph LR     A[Determinar el rendimiento E. por estilo] --&gt; B[Agrupar estudiantes por rendimiento y estilo de aprendizaje]           </pre>
Rol de tarea.	Monitoreo

# Implementación del Ciclo autonómico de Tareas.

## Ciclo 1

**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

### **Paso 2: (Análisis)**

**Tarea 4/5:** Agrupar estudiantes por rendimiento, estilo de aprendizaje actual, y hábitos en Internet

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Agrupar los alumnos de acuerdo al rendimiento obtenido por Estilo de A.	Estudiantes	Rendimiento Estilo de A.	Agrupación	Agrupar los alumnos por rendimiento y Estilo de A. utilizado

Nombre de tarea	Agrupar estudiantes por rendimiento, estilo de aprendizaje actual, y hábitos en Internet	
Descripción	Agrupar los alumnos que han tenido mejor rendimiento, en un paradigma	
Fuente de datos	Base de datos de SaCI	
Tipo de tarea de analítica de datos	Agrupación	
Técnica de analítica de datos	Series temporales	
Con que otras tareas se relaciona	 <pre> graph LR     A[Agrupar estudiantes por rendimiento y estilo de A. utilizado] --&gt; B[Definir el estilo de aprendizaje adecuado para el curso.]                     </pre>	
Rol de tarea.	Análisis	

# Implementación del Ciclo autonómico de Tareas.

## Ciclo 1

**Objetivo:** Definir el paradigma de aprendizaje adecuado para el curso que se esté dando (estudiantes y temas específicos).

**Paso 3: (Toma de Decisiones)**

**Tarea 5/5:** Determinar el nuevo estilo de aprendizaje adecuado para el curso/grupo

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Definir el estilo de A. de aprendizaje adecuado para el curso	Estudiantes	Estilo de A. Cursos Temas Estudiantes	Toma de decisiones	Determinar el estilo de A. de mayor uso con el mayor rendimiento

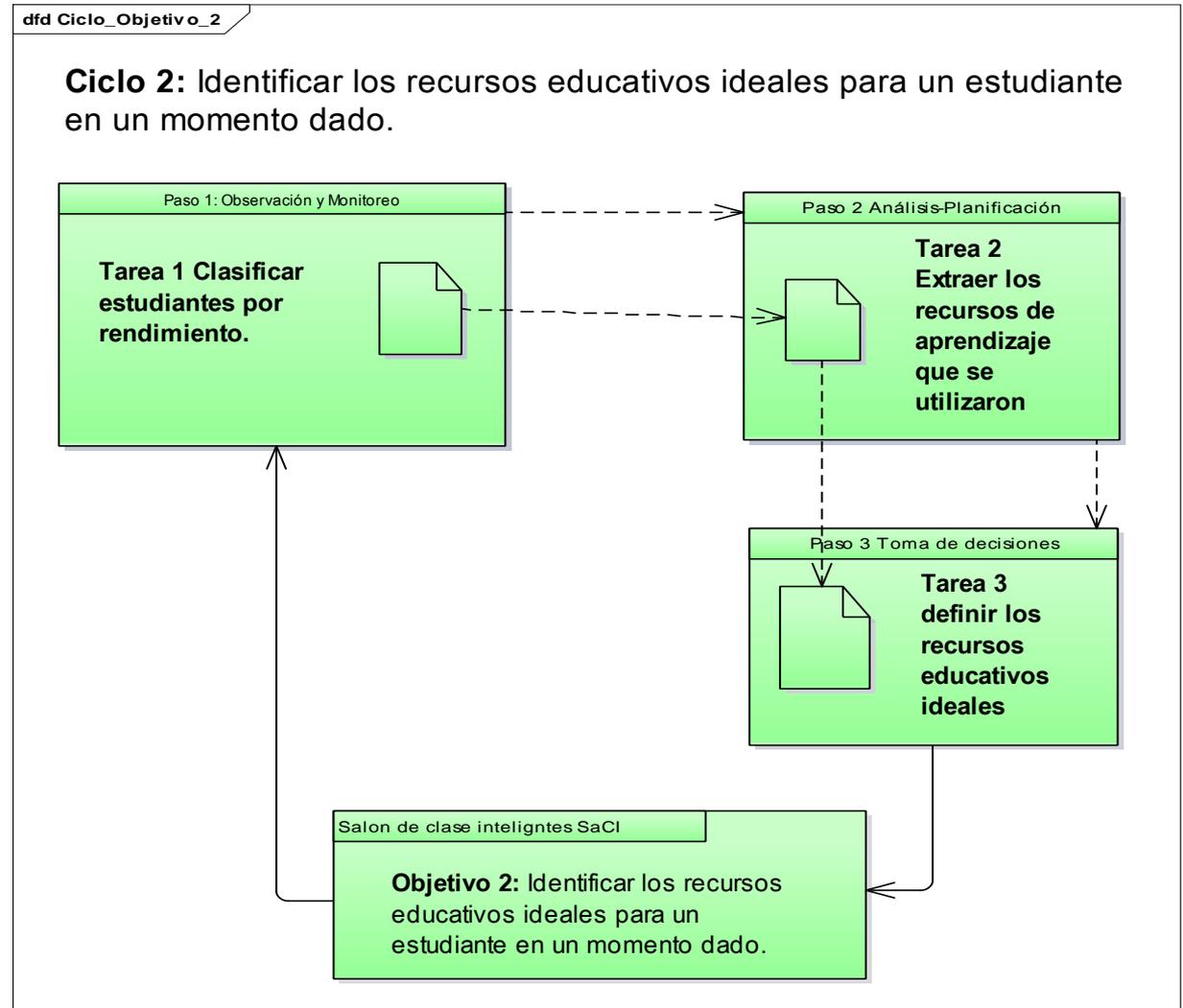
Nombre de tarea	Decidir el estilo de aprendizaje adecuado	
Descripción	Seccionar el estilo de A. que obtuvo mayor éxito en el rendimiento como el adecuado para el curso	
Fuente de datos	Base de datos de SaCI	
Tipo de tarea de analítica de datos	Toma de Decisiones	
Técnica de A.D.	Reglas de decisión	
Con que otras tareas se relaciona	<pre> graph LR     A[Definir el paradigma de aprendizaje adecuado.] --&gt; B[Tareas de SaCI]             </pre>	
Rol de tarea.	Toma de decisiones	

Ciclo autonómico de tareas para SaCI

# IMPLEMENTACIÓN DEL CICLO AUTONÓMICO DE TAREAS.

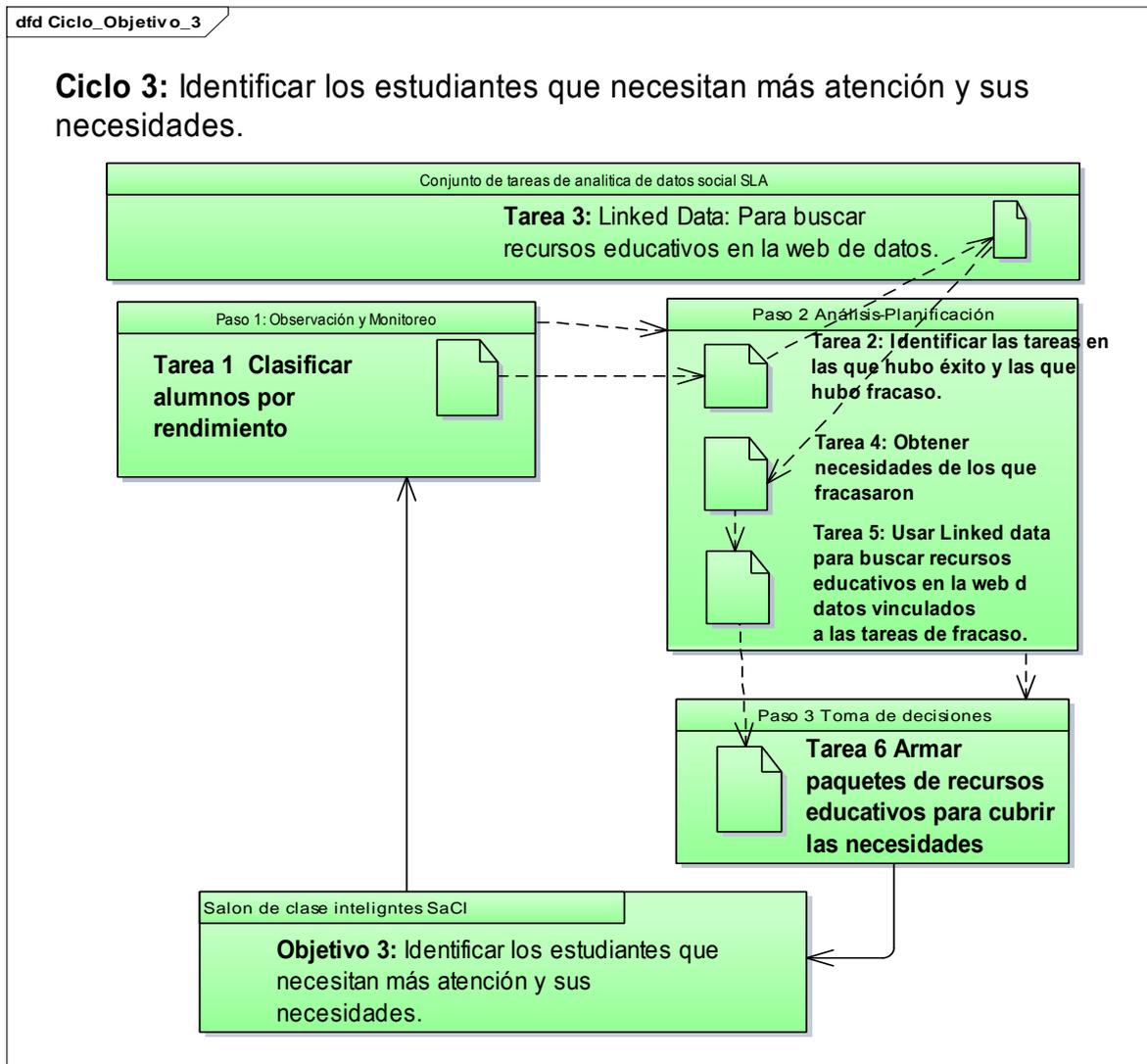
## Ciclo 2

### Modelo del ciclo 2



# IMPLEMENTACIÓN DEL CICLO AUTONÓMICO DE TAREAS. Ciclo 3

## Modelo del ciclo 3



# Implementación del Ciclo autónomo de Tareas.

## Ciclo 3

Objetivo: Identificar los estudiantes que necesitan más atención y sus necesidades.

### Paso 1: (Observación)

Tarea 1/6: Clasificar alumnos por rendimiento

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Clasificar a los alumnos por rendimiento	Estudiantes	Rendimiento Necesidades	Clasificación	clasificar estudiantes por rendimiento

Nombre de tarea	Clasificación de estudiante por rendimiento
Descripción	Tarea para clasificar a los alumnos por rendimiento para quedarse con los de bajo rendimiento
Fuente de datos	Base de Datos de SaCI
Tipo de tarea de analítica de datos	Clasificación
Técnica de analítica de datos	
Con que otras tareas se relaciona	<pre> graph LR     A[Clasificación de estudiantes por rendimiento y necesidades] --&gt; B[Determinar los rendimientos y las necesidades que requieren atención]             </pre>
Tipo de tarea en el ciclo autónomo	Monitoreo

SaCI

Ciclo autónomo de tareas para

# Implementación del Ciclo autónómico de Tareas.

## Ciclo 3

**Objetivo:** Identificar los estudiantes que necesitan más atención y sus necesidades.

**Paso 2:** (Análisis)

**Tarea 2/6:** Identificar las tareas en las que hubo éxito y las que hubo fracaso.

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Determinar Tareas exitosas y tareas con fracaso	Estudiantes	Rendimiento	Asociación Agrupación	Asociar y agrupar los alumnos para obtener detalles

Nombre de tarea	Determinar los detalles de las tareas en las que tuvo éxito y fracaso
Descripción	Agrupar los alumnos que han tenido bajos rendimiento en un curso e identificar las tareas de éxito y fracaso
Fuente de datos	Base de datos de SaCI
Tipo de tarea de analítica de datos	Asociación, Agrupación
Técnica de analítica de datos	Agrupación clustering
Con que otras tareas se relaciona	
Rol de tarea.	Análisis

# Implementación del Ciclo autónomo de Tareas.

## Ciclo 3

Objetivo: Identificar los estudiantes que necesitan más atención y sus necesidades.

**Paso 2:** (Análisis)

**Tarea 3/6:** Identificar los recursos utilizados en las tareas de fracaso.

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Recursos utilizados en las tareas de fracaso	Estudiantes	Tareas	Asociación	Asociar y agrupar los alumnos para obtener detalles

Nombre de tarea	Determinar los recursos utilizados en las tareas de fracaso
Descripción	Determinar los recursos utilizados en las tareas que han sido de fracaso dentro del grupo.
Fuente de datos	Base de datos de SaCI
Tipo de tarea de analítica de datos	Asociación, Agrupación
Técnica de analítica de datos	Agrupación clustering
Con que otras tareas se relaciona	<pre> graph LR     A[Determinar los recursos utilizados en las tareas de fracaso] --&gt; B[Obtener necesidades de las diferencias de los 2 conjuntos]             </pre>
Rol de tarea.	Análisis

# Implementación del Ciclo autónomo de Tareas.

## Ciclo 3

**Objetivo:** Identificar los estudiantes que necesitan más atención y sus necesidades.

**Paso 2:** (Análisis)

**Tarea 4/6:** Obtener necesidades de los que fracasaron

Resultados que se desean obtener	Actores	Variables asociadas	Actividades de MD a realizar	Actividades que se realizarán
Identificar las necesidades por medio de las diferencias de los conjuntos de fracaso y éxito	Estudiantes	Rendimientos Necesidades	Toma de decisiones	Asociar el rendimiento mas bajo de los estudiantes con sus necesidades

Nombre de tarea	Obtener necesidades de los que fracasaron
Descripción	Obtener necesidades de los que fracasaron
Fuente de datos	Base de datos de SaCI
Tipo de tarea de analítica de datos	Toma de decisiones
Técnica de analítica de datos	Reglas de decisión
Con que otras tareas se relaciona	<pre> graph LR     A[Obtener necesidades de las diferencias de los 2 conjuntos] --&gt; B[Fin Ciclo]             </pre>
Rol de tarea en el ciclo autónomo	Toma de decisiones

# Implementación del Ciclo autónómico de Tareas.

## Ciclo 3

**Objetivo:** Identificar los estudiantes que necesitan más atención y sus necesidades.

**Paso 2:** (Análisis)

**Tarea 5/6:** Usar Linked Data Para buscar recursos educativos en la web de datos vinculados a las tareas en las que se fracaso

Resultados Esperados	Actores	Variables	Herramienta Social	Actividades
Montar un sistema de linked Data	Temas Web de datos	Temas  Recursos educuticos.  Etiquetas	Linked Data  Modelo ontológico.  Grafos	Usar swetbok como modelo ontológico.  Definir el modelo como un grafo de temas.  Con el grafo de temas, montar un sisema de linked data para buscar recursos educativos en la web de datos

<b>Nombre de la Tarea</b>	<b>Linked Data: Para buscar recursos educativos en la web de datos</b>
---------------------------	--

<b>Descripcion</b>	Usar swebok como modelo ontológico, para definir un grafo de temas a partir de ahi, montar un sistema de linked dat para buscar recursos educuticos en la web de datos.
--------------------	---

<b>Fuente de Datos</b>	Internet
------------------------	----------

<b>Tipo de tarea de analítica de datos social</b>	Linked Data
---	-------------

<b>Técnica de analítica social</b>	Swebok como modelo ontológico
------------------------------------	-------------------------------



<b>Tipo de tarea en el ciclo</b>	Analisis
----------------------------------	----------

Ciclo autónómico de tareas para SaCI

# Implementación del Ciclo autónómico de Tareas.

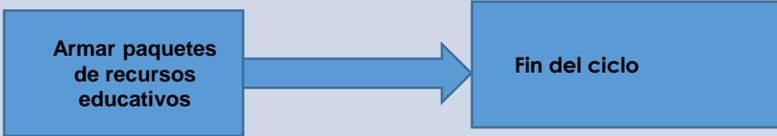
## Ciclo 3

**Objetivo:** Identificar los estudiantes que necesitan más atención y sus necesidades.

**Paso 3:** (Toma de decisiones)

**Tarea 6/6:** Armar paquete de recursos educativos para cubrir las necesidades

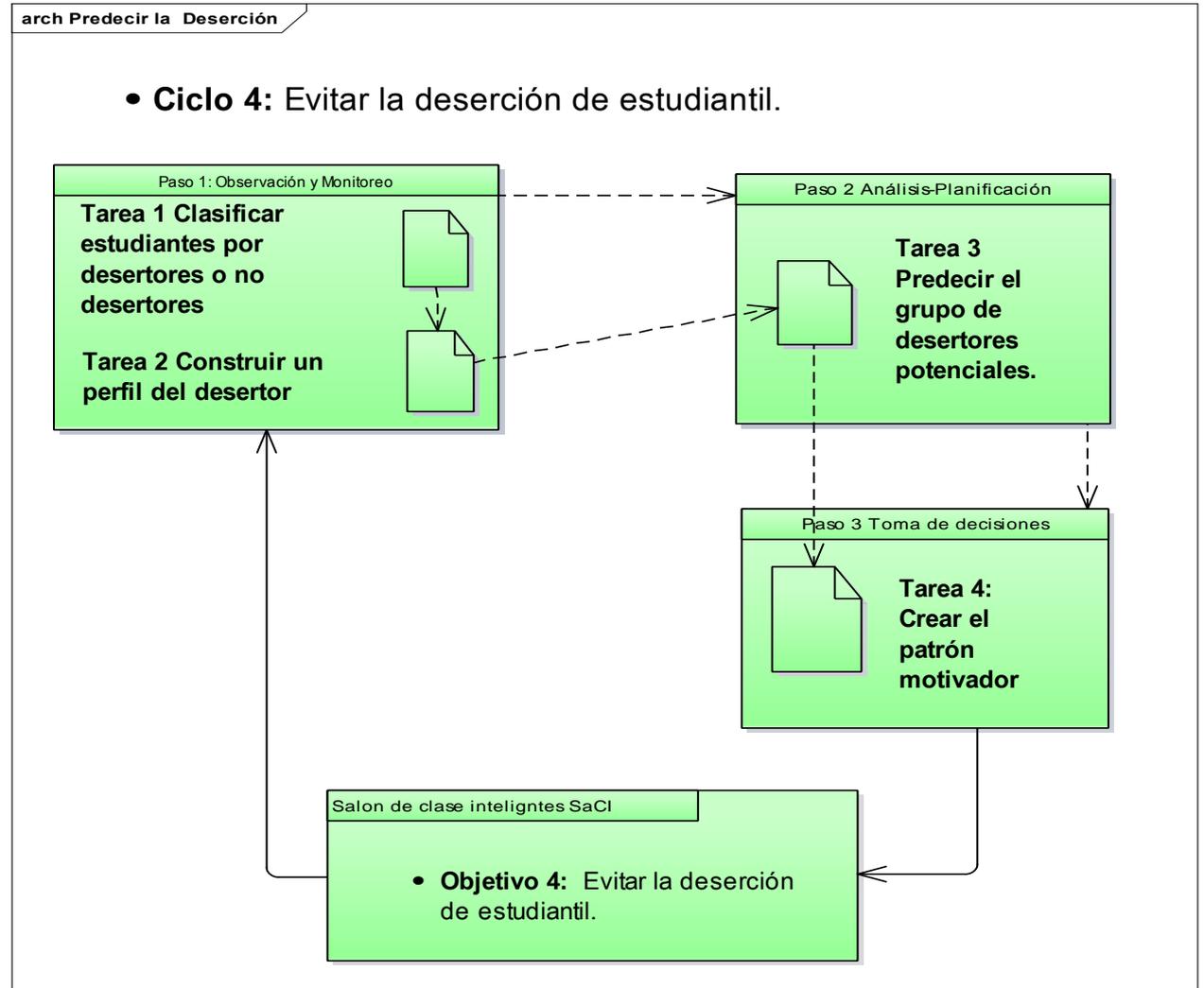
Resultados Esperados	Actores	Variables	Herramienta Social	Actividades
Armar paquetes de recursos educativos para cubrir necesidades	Temas Web de datos	Temas  Recursos educativos.  Etiquetas	Linked Data  Modelo ontológico.  Grafos	Armar paquetes de recursos educativos

Nombre de la Tarea	Armar paquetes de recursos educativos para cubrir necesidades
Descripción	Armar paquetes de recursos educativos para cubrir necesidades de los estudiantes.
Fuente de Datos	Internet
Tipo de tarea de analítica de datos social	Linked Data
Técnica de analítica social	Swebok como modelo ontológico
Tarea con la que se relaciona	 <pre> graph LR     A[Armar paquetes de recursos educativos] --&gt; B[Fin del ciclo]             </pre>
Tipo de tarea en el ciclo	Analisis

# IMPLEMENTACIÓN DEL CICLO AUTONÓMICO DE TAREAS.

## Ciclo 4

### Modelo del ciclo 4



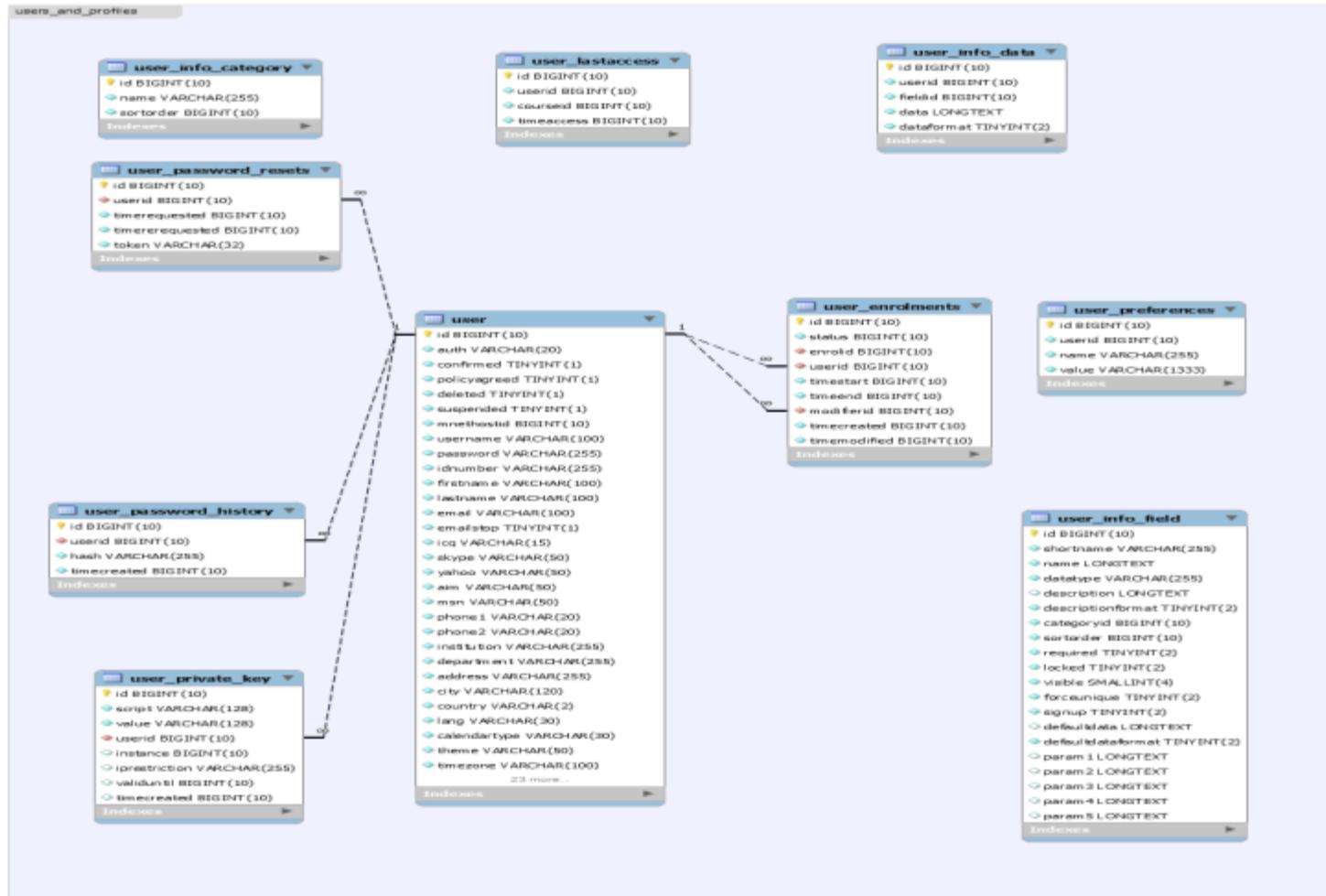
# Modelo de datos de SaCI

## Fuentes

- **Fuentes de datos Virtual Learning Environment (VLE):** Son las fuentes de datos provenientes de sistemas automáticos de enseñanza como por ejemplo los moodle.
- **Fuentes de Datos de Sistemas de Gestión académica (SGA):** Son fuentes de datos provenientes de sistemas internos para la gestión de notas de los estudiantes, por ejemplo OpenSIS.
- **Fuentes de datos de la nube:** Son todas las fuentes de datos disponibles en internet y de repositorios de datos como por ejemplo las redes sociales, wikidata etc.

# Modelo de datos de SaCI Ciclo 1

Modelo de datos VLE para los usuarios



Ciclo autónomico de tareas para SaCI



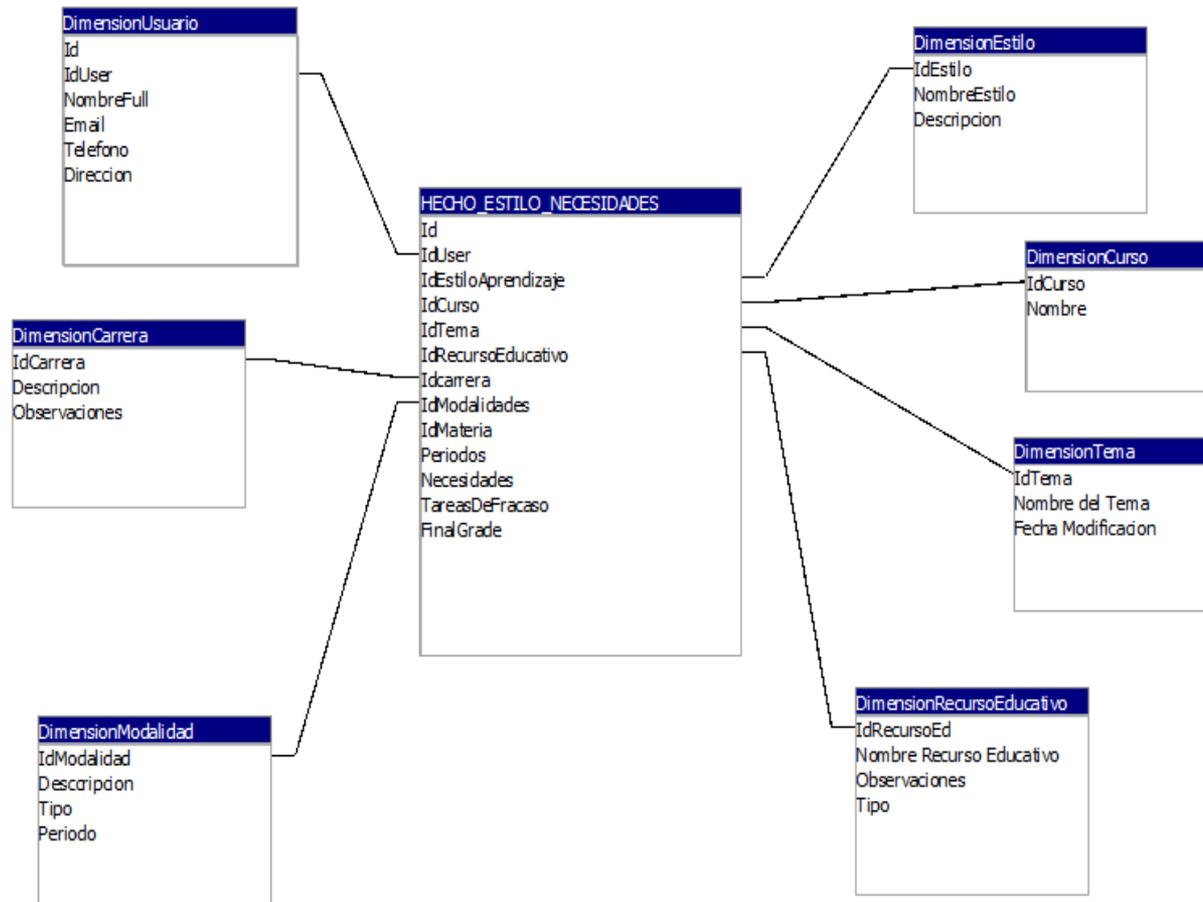
# Modelo de datos de SaCI Ciclo 1

## Tablas de VLE utilizadas para la muestra de datos

Tabla	Descripción
mdl_log	Registra de eventos LOG del VLE
mdl_user	Registra a todos los usuarios del VLE
mdl_grade_grades finalgrade	Almacena la nota final de un curso en el VLE
mdl_quiz	Registra cada uno de los exámenes del VLE
mdl_lessons	Registra cada una de las lecciones del VLE
mdl_workshop	Registra los trabajos y asignaciones en el VLE
mdl_quiz_grades	Registra loss resultados de los exámenes
mdl_course	Registra los cursos VLE

# Modelo de datos de SaCI Ciclo 1

## Modelo Multidimensional para el Ciclo 1



# Operaciones Extraccion Transformacion y carga ETL

**Objetivo 1:** Definir el estilo de aprendizaje adecuado para el curso que se esté dando  
**Tarea 1:** Clasificar los estudiantes de un curso por estilo de aprendizaje

La Tabla que se presenta a continuación, contiene las variables de relevancia para el objetivo 1 y son tomadas de las tablas seleccionadas anteriormente.

Tarea de Analítica de Datos	Variable	Descripción	Tipo	Fuente
Tarea 1	userid	Identificación única para el agente en el sistema	Bigint(10)	mdl_log
Tarea 1	Idcourse	Identificación única para el curso	Bigint(10)	mdl_log
Tarea 1	estiloAprendizaje	Identifica el estilo de aprendizaje utilizado por el estudiante	String	Fuente externa
Tarea 1	Cmid	Identificación del Módulo donde se realizó la acción	Bigint(10)	mdl_log
Tarea 1	fullName	Nombre completo del estudiante	String	mdl_log
Tarea 1	action	Identifica la acción que realiza el usuario en la plataforma	Varchar	mdl_log
Tarea 1	time	Fecha y hora en que se produce el registro en la tabla log	bitint	mdl_log

**Tarea 3** agrupar estudiantes por rendimiento y estilo de aprendizaje utilizado

Tarea de Analítica de Datos	Variable	Descripción	Tipo	Fuente
Tarea 3	userid	Identificación única para el agente en el sistema	Bigint(10)	mdl_log
Tarea 3	Idcourse	Identificación única para el curso	Bigint(10)	mdl_log
Tarea 3	Grade	Almacena las calificaciones de los alumnos para obtener el rendimiento	Number(10,5)	quiz_grades

**Tarea 2:** Determinar el rendimiento de los estudiantes por estilo de aprendizaje.

Tarea de Analítica de Datos	Variable	Descripción	Tipo	Fuente
Tarea 2	userid	Identificación única para el agente en el sistema	Bigint(10)	mdl_log
Tarea 2	Idcourse	Identificación única para el curso	Bigint(10)	mdl_log
Tarea 2	grade	Almacena las calificaciones de los alumnos	Number(10,5)	quiz_grades
Tarea 2	cmid	Identificación del Módulo donde se realizó la acción	Bigint(10)	mdl_log
Tarea 2	fullName	Nombre completo del estudiante	String	mdl_log
Tarea 2	action	Identifica la acción que realiza el usuario en la plataforma	Varchar	mdl_log
Tarea 2	time	Fecha y hora en que se produce el registro en la tabla log	bitint	mdl_log

**Tarea 4:** definir el estilo de aprendizaje adecuado para el curso

Tarea de Analítica de Datos	Variable	Descripción	Tipo	Fuente
Tarea 4	userid	Identificación única para el agente en el sistema	Bigint(10)	mdl_log
Tarea 4	Idcourse	Identificación única para el curso	Bigint(10)	mdl_log
Tarea 4	Grade	Almacena las calificaciones de los alumnos para obtener el rendimiento más exitoso	Number(10,5)	quiz_grades
Tarea 4	Estilo de aprendizaje	Estilo de aprendizaje		Fuente externa

## Operaciones Extracción Transformación y carga ETL

### Operaciones Extracción Transformación y carga

Tarea que la consume	Variable	Fuente-Extracción	Proceso	Carga
Tarea 1 Tarea 2 Tarea 3 Tarea 4	<u>userid</u>	<u>mdl_user</u>	No sufre cambios	Se carga en la tabla de hecho
Tarea 1 Tarea 2 Tarea 3 Tarea 4	<u>courseid</u>	<u>mdl_log</u>	No sufre cambios	Se carga en la tabla de hecho
Tarea 1	<u>cmid</u>	<u>mdl_log</u>	No sufre cambios No cargar Nulos	Se carga en la tabla de hecho
Tarea 1 Tarea 2	<u>fullName</u>	<u>mdl_log</u>	No sufre cambios Se omiten nulos	Se carga en la tabla dimensional de alumnos
Tarea 1	<u>action</u>	<u>mdl_log</u>	No sufre cambios Se omiten nulos	Las acciones pueden ser:
Tarea 1	<u>time</u>	<u>mdl_log</u>	No sufre cambios Se omiten nulos	Se carga en la tabla de hecho
Tarea 2	<u>grade</u>	<u>mdl_grade_grades</u>	No sufre cambios Se omiten nulos	Se almacena en la tabla de hecho
Tarea 1	<u>nombreEstilo</u>	Fuente externa	Se calcula con valores de <u>mdl_log</u>	Se carga en la tabla dimensional de estilos
	<u>email</u>	<u>mdl_user</u>	No sufre cambios	Se carga en la tabla dimensional de estilos
	Teléfono	<u>mdl_user</u>	No sufre cambios	Se carga en la tabla dimensional de usuarios
Ciclo 4 Tarea 3	Materia	<u>Mdl_course</u>	No sufre cambios	Se carga en la tabla dimensional de materias