



UNIVERSIDAD DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA QUÍMICA, BIOTECNOLOGÍA Y
MATERIALES

***Smart Training*, Servicio web de Entrenamiento de**
Modelos

DAVID ALFREDO MEDINA ORTIZ

Profesor Tutor: **ÁLVARO OLIVERA NAPPA, PH.D.**

Resumen Desarrollo de Proyecto,
Seminario 02

Santiago – Chile
Agosto, 2018

TABLA DE CONTENIDOS

página

Tabla de Contenidos

Índice de Figuras	II
Índice de Tablas	III
Resumen	IV
1. INTRODUCCIÓN	5
1.1. Data Mining	5
1.2. Problemática	6
1.2.1. Estado del Arte	7
1.3. Smart Training	7
1.3.1. Módulo de Procesamiento de Datos	8
1.3.2. Módulo de Análisis Estadístico	8
1.3.3. Módulo de Análisis de Features	8
1.3.4. Módulo de Clustering	8
1.3.5. Módulo de Entrenamiento de Modelos	9
1.3.6. Diagrama solución	9
1.4. Objetivos y Alcances	9
1.5. Metodología de Desarrollo de Software	10
2. Análisis	12
2.1. Funciones del Sistema	12
2.1.1. Funciones asociadas al Usuario	13
2.1.2. Funciones asociadas a las acciones en el Acceso al Sistema	17
2.1.3. Funciones asociadas a las estadísticas de uso	20
2.1.4. Funciones asociadas al módulo de estadísticas en el set de datos	21
2.1.5. Funciones asociadas al set de datos	26
2.1.6. Funciones asociadas al módulo de algoritmos de aprendizaje no supervisado	28
2.1.7. Funciones asociadas al módulo de clasificaciones	33
2.1.8. Funciones asociadas al módulo de análisis de características	40
2.1.9. Funciones asociada al Sistema de Colas	43

2.1.10. Funciones asociadas al Sistema de Notificación	46
2.1.11. Funciones asociadas al Sistema de Almacenamiento Persistente . . .	47
2.2. Atributos del Sistema	49
2.3. Actores y Usuarios	58
2.4. Casos de Uso	60
2.4.1. Casos de uso asociados a las acciones del Usuario administrador . .	62
2.4.2. Casos de uso asociados al uso de recursos	84
2.4.3. Casos de uso asociados al sistema de almacenamiento persistente . .	93
2.4.4. Casos de uso asociados al sistema de notificación	108
2.4.5. Casos de uso asociados al manejo de sesiones	111
2.4.6. Casos de uso asociados al módulo estadísticas de uso	127
2.4.7. Casos de uso asociados a la generación de reportes	140
2.4.8. Casos de uso relacionados al módulo de análisis estadístico	144
2.4.9. Casos de uso asociados al módulo de revisión de los datos de entrada	165
2.4.10. Casos de uso asociados al módulo de clustering	168
2.4.11. Casos de uso relacionados al lanzamiento de Jobs	196
2.4.12. Diagramas de casos de uso	212
2.5. Diagramas de secuencia o colaboración	212
2.6. Conceptos	212
2.6.1. Modelo Conceptual	212
2.7. Entidades	212
2.7.1. Modelo de Entidades	212
3. Diseño	213
3.1. Arquitectura de Software	213
3.2. Diagramas de Interacción	213
3.3. Diagrama de Clases	213
3.4. Diagramas de Estado	213
4. Planificación	214
4.1. Etapas del Proyecto	214
Bibliografía	215

ÍNDICE DE FIGURAS

	página
1.1. Componentes en la minería de datos	5

ÍNDICE DE TABLAS

	página
2.1. Resumen categorías de funciones.	13
2.2. Funciones relacionadas a las acciones de usuario	17
2.3. Funciones asociadas a la sección de Acceso al sistema	20
2.4. Funciones asociadas a las estadísticas de uso	21
2.5. Funciones asociadas al módulo estadístico.	26
2.6. Funciones asociadas al set de datos.	28
2.7. Funciones asociadas a los módulos de clustering	33
2.8. Funciones asociadas al módulo de clasificación.	40
2.9. Funciones asociadas al módulo de análisis de características	43
2.10. Funciones asociadas al Sistema de Colas	46
2.11. Funciones asociadas al Sistema de Notificaciones	47
2.12. Funciones asociadas al Sistema de Almacenamiento Persistente	49
2.13. Atributos presentes en el sistema	50
2.14. Tabla resumen de Atributos por Función	58
2.15. Actores que interactúan con el sistema	60
2.16. Perfiles de usuario identificados	61
2.17. Casos de uso asociados a los usuarios administrativo y root	84
2.18. Casos de uso asociados al uso de recursos.	93
2.19. Casos de uso relacionados al sistema de almacenamiento persistente.	108
2.20. Casos de uso relacionados al sistema de notificaciones	111
2.21. Casos de uso asociados al manejo de sesiones en el sistema	127
2.22. Casos de uso asociados al módulo estadístico	140
2.23. Casos de uso asociados a reportes	144
2.24. Casos de uso relacionados al módulo de análisis estadístico	165
2.25. Casos de uso relacionados a la revisión del set de datos	168
2.26. Casos de uso relacionados a los módulos de clustering	196
2.27. Casos de uso asociados al lanzamiento de Jobs	211

RESUMEN

Conceptos como minería de datos, machine learning, big data, análisis estadísticos, modelamientos matemáticos, etc, son mencionados día a día, ya sea en el ámbito privado como público, involucrando áreas como: comercio, salud, investigación, transporte, etc. Lo cual denota que son temáticas que han adquirido mayor relevancia y su ascenso seguirá con el pasar del tiempo.

La manipulación de grandes volúmenes de datos, con el fin de poder extraer información de ellos, búsqueda de patrones, evaluaciones estadísticas, etc. Implica por parte del interesado, tener conocimientos en dichas áreas además de comprender herramientas informáticas que le permitan dicho procedimiento. Sin embargo, dichas herramientas o son costosas, debido a la licencia que implica, o, se requiere de conocimiento informático para su manipulación, debido a que requiere implementar módulos o servicios a medida que permitan ejecutar las tareas de interés, lo cual deja a un número importante de entidades que desean involucrarse en dicho mundo, pero no cuentan con las capacidades ni tampoco con las competencias para ello.

Dado a lo anterior y en base a la creciente demanda de desarrollo de metodologías que permitan aplicar data mining a procesos de datos, con el fin de extraer información y conocimiento de la misma, se propone Smart Training, como sistema web, que facilite los procesos de evaluaciones estadísticas, búsqueda de patrones de comportamiento, desarrollo de modelos de clasificación y evaluación de características o features en el set de datos a estudiar.

1. INTRODUCCIÓN

1.1. Data Mining

Minería de datos es el proceso de descubrimiento de patrones en set de datos, involucrando métodos asociados a Machine Learning, Estadísticas y sistemas de bases de datos. [1]. La minería de datos es un subcampo interdisciplinario de la informática, el cual tiene por objetivo general extraer información (a través de métodos inteligentes) de un conjunto de datos y transformar la información en una estructura comprensible para su uso posterior. [2, 3]. La minería de datos es el paso de análisis del proceso de *descubrimiento de conocimiento en bases de datos*, o KDD. [4]. Además del análisis en bruto de los datos, también incluye aspectos de manipulación de bases de datos, pre procesamiento de datos, evaluaciones de modelo e inferencia, métricas de interés, consideraciones de complejidad, post procesamiento de estructuras descubiertas, visualización y actualización de la información.

En la Figura 1.1, se exponen las principales ramas que componen la minería de datos y los diferentes procesos que se asocian a dichas ramas.

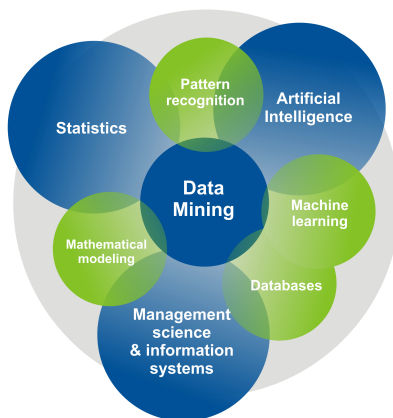


Figura 1.1: Componentes en la minería de datos

Son tres las principales áreas que abarca la minería de datos: Estadística, Inteligencia Artificial y Manipulación de sistemas de información, mientras que son distintos procesos los que interactúan entre estas ramas, tales como: Modelamiento Matemático, reconocimiento de patrones, Sistemas de almacenamiento persistente y machine learning.

Cada área en particular tiene un objetivo general y diversos objetivos específicos. Sin embargo, estas áreas interactúan entre sí, con el fin de poder extraer patrones de información que generen conocimientos a partir de la data de procesada.

La minería de datos se utiliza en diferentes campos, tales como: Genética, Evaluaciones proteómicas, Comercio, Sistemas de tránsito, Optimizaciones en procesos industriales, reconocimiento de patrones y rasgos cuantificables en enfermedades y más recientemente en áreas de dinámicas moleculares y parámetros para la generación de pipe lines automatizados de simulaciones cuánticas en sistemas químicos.

El uso de la minería de datos y la búsqueda de patrones de comportamientos de datos de interés, no sólo es un área que se enfoca en la investigación. Diversas son las entidades privadas que ofrecen servicios de big data y data science, además del sector público, con el fin de detectar puntos críticos de zonas de riesgo, zonas de accidentes, evaluación o perfilación de grupos de estudio, etc.

1.2. Problemática

Actualmente, los sistemas de almacenamiento persistente, permiten disponer de información que es de interés para distintos tipos de entidades, las cuales dependerán del área a las que se dediquen. Sin embargo, cada área en particular, tiene objetivos similares, tales como:

- Qué significa los datos que tengo?
- Puedo personalizar y interpretar datos?
- Puedo optimizar procesos en base a la información que tengo sobre estos?
- Es factible aumentar la experiencia de usuario en cuanto a procesos de ventas y compras, personalizando sus áreas de cliente?

Son muchos los objetivos que se pueden encontrar y muchas las aplicaciones que implica esta técnica. No obstante, el hecho de tener data almacenada y no saber cómo procesarla es un gran problema para muchas pequeñas y medianas empresas, así como para también ventas, laboratorios de investigación, etc.

El deseo de aplicar minería de datos, con el fin de extraer información, es día a día, más frecuente. Sin embargo, un usuario común debe enfrentar la problemática de como afrontar el problema, adquirir las competencias, o simplemente, contratar servicios de data science, los cuales cobran altas sumas de dinero y emplean un tiempo elevado con el fin de llegar a una respuesta pronta.

Por otro lado, existen herramientas que facilitan la utilización de minería de datos, pero, el costo por conceptos de licencia es demasiado elevado y no permite abarcar diversas áreas y testear algoritmos variados. Mientras que por otro lado, existen módulos o librerías que han sido implementadas en diversos lenguajes de programación, con el fin de aplicar minerías de datos, pero esto aumente aún más la complejidad del tema, debido a que para implementar dicha labor, se requiere de conocimientos en programación y en algunos casos, las implementaciones son engorrosas y requieren de un conjunto de arquitecturas que soporten dichas instancias.

Dado a lo anterior, es que se propone el desarrollo de una herramienta web, que permita la aplicación de diversas técnicas de minería de datos y oriente de manera inteligente al usuario, esto implicaría que no se requiere de un conocimiento de programación, además que las competencias en minería de datos deben ser mínimas, puesto que la idea contempla la orientación al usuario con respecto al objetivo que desea.

Durante este documento, se expone el diseño de la herramienta, las metodologías a utilizar y los artefactos de software que se crearán, con el fin de poder implementar esta herramienta en base a una metodología de diseño, además se expone un resumen de las tecnologías actuales, cuales son las ventajas y desventajas que poseen cada una y en qué se diferencia el software planteado con respecto a los existentes.

1.2.1. Estado del Arte

1.3. Smart Training

Smart Training es un sistema web, que facilita la utilización de técnicas de minería de datos, con el fin de evaluar características en set de datos, reconocer patrones, entrenar algoritmos de clasificación, generar evaluación de características, etc. Tiene por finalidad acercar la minería de datos a un público que no posee las competencias para implementar modelos mediante utilización de módulos de programación.

Smart Training se compone de 5 módulos, los cuales se detallan a continuación.

1.3.1. Módulo de Procesamiento de Datos

Este módulo tiene por objetivo cargar la data entregada en archivos de texto, evalúa los datos existentes, corrobora que no existan problemas, revisa el set de datos, encuentra valores nulos, codificaciones no permitidas, etc, a lo que, finalmente, entrega un resumen del proceso, exponiendo los resultados y de dicha tarea y comentando si es posible trabajar con dicho set cargado, en caso contrario, expone mensajes con recomendaciones a seguir.

1.3.2. Módulo de Análisis Estadístico

Este módulo permite la evaluación del set de datos contemplando, correlaciones, box plot, distribuciones mediante histogramas, scatter plot, gráficos de densidad de datos, además de resúmenes estadísticos para cada atributo o característica en el set de datos que se está trabajando.

1.3.3. Módulo de Análisis de Features

Este módulo permite evaluar las características en el set de datos y el aporte que éstas entregan, adicional a ello, contempla análisis mediante técnicas PCA, para generar reducciones de dimensionalidad, Mutual Information y técnicas basadas en correlación, todas con el objetivo de explicar los comportamientos de éstas y cómo influyen en el set de datos.

1.3.4. Módulo de Clustering

Clustering es una de las técnicas más conocidas para asociar segmentos en una muestra, es decir, generar grupos o particiones que tengan un alto grado de diferencia entre ellas, pero cuyos integrantes en una partición dada, sean altamente similares.

Existen diferentes algoritmos de clustering y parámetros asociados a estos, los cuales tienen formas de encontrar particiones distintas, basándose en distancias, medidas gaussianas, generación de hiper planos, etc.

Este módulo tiene por objetivo generar exploración de dichas técnicas y algoritmos, con el fin de poder entregar particiones en el set de datos, además, permite la evaluación de dichas particiones con el fin de poder determinar si son estadísticamente significativas o no, además de cumplir con los criterios de similitud y diferenciación mencionados previamente.

Normalmente, la búsqueda de particiones conlleva al hecho de generar modelos de clasificación para nuevos ejemplos y determinar a qué particiones se encuentran, o también, generar divisiones para implementar set de datos diferentes con comportamientos

diferentes, de tal manera que a la hora de aplicar algoritmos de clasificación, sus comportamientos presenten mayor eficiencia.

1.3.5. Módulo de Entrenamiento de Modelos

Entrenar un modelo de clasificación, predicción, implica tener un conjunto de elementos con su clasificación o valor de predicción conocido, con el fin de poder, a partir de éste, evaluar nuevos ejemplos, ya sea para clasificarlos o para predecir posibles valores de interés. Todas estas tareas, aplicando minería de datos, se logran mediante la implementación de algoritmos de aprendizaje supervisado.

Existen diferentes algoritmos de aprendizaje supervisado, los cuales contemplan diferentes formulaciones matemáticas y características, los cuales cumplen con dicha tarea, cada uno de estos, presenta características distintas, en relación al funcionamiento del mismo, la elección de un algoritmo por sobre otro, va de la mano con el hecho de las necesidades que el problema conlleva, ya sea, con el fin de entregar sólo un resultado, adicionar valores que permitan explicar el porqué de la clasificación, etc. Normalmente, para un problema desconocido, es necesario implementar fases exploratorias que permitan evaluar diferentes algoritmos y sus parámetros. Con el fin de poder, a partir de dicha instancia y en base a métricas que permitan evaluar el desempeño, seleccionar un algoritmo y sus parámetros.

Lo anterior, es el objetivo del módulo de entrenamiento de modelos, la idea de éste, es recibir un set de datos con ejemplos clasificados o cuya respuesta tenga un valor conocido y aplicarle diversos algoritmos y variaciones de parámetros, entregando un resumen de las medidas de desempeño obtenidas y efectuando un ranking según medida, para que finalmente se pueda entregar una recomendación de los mejores modelos para un cierto problema.

1.3.6. Diagrama solución

1.4. Objetivos y Alcances

El objetivo general del proyecto contempla la implementación de un sistema web denominado, *Smart Training*, el cual permita la aplicación de distintas técnicas de minería de datos a set de datos de interés del usuario, a partir de los cuales éste pueda entender comportamientos de datos y generar conocimiento a partir de ellos.

Es importante destacar los objetivos específicos que nacen dentro del desarrollo del proyecto.

1. Diseñar metodología de software a utilizar.
2. Comprender los requerimientos observados a partir del estado del arte.
3. Evaluar las funcionalidades y atributos que tendrá el sistema.
4. Comprender los usuarios y actores del software.
5. Entender las secuencias y flujos de trabajo existentes en la herramienta.
6. Crear modelos de conceptos, entidades y clases.
7. Implementar los módulos propuestos.
8. Implantar sistema.

1.5. Metodología de Desarrollo de Software

Existen diversas metodologías de desarrollo de software, las cuales contemplan diferentes características y se enfocan en distintos puntos objetivo. Algunos ejemplos son.

- Metodologías Ágiles.
- Diseño cascada.
- Diseño estrella.
- Iterativas.

Metodología ágil, se utiliza cuando el objetivo se basa principalmente en sacar a producción el software de manera rápida, no contempla procesos de diseño complejos y simplemente se centra en el desarrollo del producto, lo cual permite, por un lado, poseer un producto en poco tiempo. Sin embargo, está sometida a enmarcar errores debido a que no se consideraron patrones de diseño.

El diseño en cascada y también en estrella, se centran en los requerimientos de usuario y generar sub productos asociados al software final, los cuales cumplen un objetivo en particular del software, fueron muy utilizadas en los años 90, debido a la simplicidad que estos poseían. No obstante, no contempla iteraciones para evaluar los flujos de trabajo, ni tampoco la utilización de paradigmas complejos de desarrollo de software.

Una de las metodologías más utilizadas, es la Iterativa, ésta contempla un conjunto de patrones de diseño, los cuales están asociados a la evaluación de las funcionalidades, los atributos, describir secuencias de flujos y asociar conceptos, es la metodología que

contempla un mayor conjunto de pasos. No obstante, es la que más asegura que a la hora de implementar, dicho proceso sea rápido. El hecho de ser iterativa, implica que cada etapa entrega un artefacto de software, bajo el cual depende el siguiente, en nuevas etapas se hacen mejoras a los artefactos generados y se está en constante feedback.

Adicional a las metodologías de software, existen diferentes paradigmas de programación que son empleados en conjunto con dichas estrategias. Los dos principales son: Estructurado y Orientado a Objetos. En el primero, se sigue un orden secuencial del problema a desarrollar, no está adaptado para grandes desarrollos, si no que más bien, es empleado en scripting y manejo de patrones en archivos de texto. Por otro lado, la programación Orientada a Objetos (POO), cumple con la características de ser más cercana a la vida real, debido a que se basa en el diseño de clases, que representan entidades las cuales pueden ser ficticias o reales. Presenta grandes ventajas debido a que posee las siguientes características.

- *Encapsulamiento*: Un Objeto es dueño de sus atributos y métodos.
- *Polimorfismo*: Un mismo método, pueden tener significados diferentes para distintas clases.
- *Reusabilidad*: Una clase modelada puede utilizarse en diversos proyectos debido a que siempre poseerá los mismos atributos y métodos.

Además de dichas características, la POO asocia conceptos abstractos a la programación, tales como: herencia, dependencias, asociaciones y composiciones, las cuales aumentan más aún, la usabilidad de este paradigma.

En esta ocasión, debido a la envergadura del proyecto, a las características que se espera que posea y a las ventajas que entregan las metodologías, se utilizará el diseño de software iterativo con objeto acoplado. Es decir, se diseñará teniendo en consideración distintas iteraciones asociadas a los artefactos de software que se desarrollen, enfocando dicho diseño a POO.

2. Análisis

La etapa de análisis del proyecto, contempla la evaluación y determinación de las diferentes funcionalidades que tendrá el sistema, asociado a los atributos que estos presentan y que permiten cuantificar de cierta manera el software. También contempla las evaluaciones de los flujos asociados a cada función y determina las secuencias de pasos a seguir para dar respuesta a cada una de éstas, exponiéndolos en forma narrativa mediante los casos de uso. Finalmente se evalúan los conceptos existentes y que representan entidades dentro del software, los cuales forman parte del diseño posterior.

2.1. Funciones del Sistema

Las funciones del sistema, hacen referencia a las acciones que se pueden desarrollar en software, las cuales representan flujos completos de trabajo en cuanto a nivel de interacción y coexisten al menos un actor y un concepto.

A continuación se listan un conjunto de funciones, las cuales han sido divididas con respecto a la finalidad que éstas cumplen en el software, cada función posee su número de registro, además de presentar su categoría y su sigla asociada, esto es con el objetivo de poder asociar funciones a módulos y agruparlas según el objetivo o la relación que éstas tengan.

Un resumen de estas funciones se expone en la Tabla 2.1.

A continuación, se lista y describen cada funciones, según la categoría a la que pertenecen.

<i>Descripción de categoría de funciones</i>			
#	ID	Nombre Categoría	Descripción
1	US	Usuario	Funciones relacionadas a las acciones a realizar por parte del usuario
2	AC	Acceso	Funciones relacionadas al sistema de acceso a la aplicación
3	ES	Estadísticas	Funciones relacionadas a las estadísticas de uso de la herramienta
4	AE	Análisis estadístico	Funciones relacionadas al análisis estadístico de los datos de entrada
5	ESD	Evaluación set de datos	Funciones relacionadas al chequeo y validación de los datos de entrada
6	MANS	Modelos de Aprendizaje No Supervisado	Funciones relacionadas al uso de clustering
7	MAS	Modelos de Aprendizaje Supervisado	Funciones relacionadas al entrenamiento de modelos
8	AF	Análisis de Características	Funciones relacionadas al análisis de características
9	SC	Sistema de Colas	Funciones relacionadas al proceso de jobs en sistemas de colas
10	SN	Sistema de notificaciones	Funciones relacionadas a las notificaciones vía correo electrónico
11	AP	Almacenamiento Persistente	Funciones relacionadas al sistema de almacenamiento persistente

Tabla 2.1: Resumen categorías de funciones.

2.1.1. Funciones asociadas al Usuario

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción

1	US	Usuario	Registrar nuevo usuario	El sistema deberá permitir el registro de un nuevo usuario en el almacenamiento persistente
2			Editar usuario existente	El sistema deberá permitir la edición de un usuario existente
3			Eliminar usuario existente	El sistema deberá permitir la eliminación de un usuario existente

4		Visualizar usuario	El sistema deberá permitir listar a todos los usuarios y un detalle de los mismos
5		Crear áreas de trabajo	El sistema deberá permitir crear áreas de trabajo para un nuevo usuario
6		Eliminar áreas de trabajo	El sistema deberá permitir eliminar áreas de trabajo para un usuario existente

7		Registrar roles	El sistema deberá permitir el registro de nuevos roles en el almacenamiento persistente
8		Editar roles existente	El sistema deberá permitir la edición de un rol existente
9		Eliminar rol existente	El sistema deberá permitir la eliminación de un rol existente

10			Visualizar roles	El sistema deberá permitir listar los roles existentes en el almacenamiento persistente
----	--	--	------------------	--------------------------------------------------------------------------------------------------------------

Tabla 2.2: Funciones relacionadas a las acciones de usuario

2.1.2. Funciones asociadas a las acciones en el Acceso al Sistema

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción
1			Iniciar sesión de usuario	El sistema deberá permitir iniciar sesión en el sistema

2			Autenticar usuario	El sistema deberá permitir comprobar las credenciales del usuario para el inicio de sesión
3			Cerrar sesión	El sistema deberá permitir cerrar la sesión actual del usuario
4			Recuperar cuenta de usuario	El sistema deberá permitir recuperar la cuenta de usuario en caso de olvidar la contraseña

5			Reestablecer cuenta de usuario	El sistema deberá permitir reestablecer la cuenta de usuario si se solicita una recuperación
6			Modificar datos de acceso	El sistema deberá permitir modificar los datos de acceso asociados al inicio de sesión
7			Notificación de cambios	El sistema deberá permitir notificar el cambio de datos en las credenciales generado

Tabla 2.3: Funciones asociadas a la sección de Acceso al sistema

2.1.3. Funciones asociadas a las estadísticas de uso

Funciones del Sistema				
#	ID Categoría	Categoría	Función	Descripción
1	ES	Estadísticas	Visualizar estadísticas de uso	El sistema deberá permitir entregar estadísticas de uso por usuario y por módulo
2			Visualizar carga de servidores	El sistema deberá permitir visualizar la carga actual de los servidores

3			Visualizar carga de sistema de colas	El sistema deberá permitir visualizar la carga del sistema de colas
4			Visualizar estados de trabajos	El sistema deberá permitir visualizar los estados de trabajo
5			Visualizar estadísticas de trabajos	El sistema deberá permitir visualizar estadísticas relacionadas a los trabajos procesados

Tabla 2.4: Funciones asociadas a las estadísticas de uso

2.1.4. Funciones asociadas al módulo de estadísticas en el set de datos

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción
1	AE	Análisis estadístico	Estimar estadísticos en relación a la muestra	El sistema deberá permitir calcular estadísticos sobre la data entregada
2			Visualizar estadísticos de datos	El sistema deberá permitir visualizar estadísticos sobre la data entregada
3			Estimar box plot	El sistema deberá permitir estimar el box plot para un set de datos dado

4		Visualizar box plot	El sistema deberá permitir visualizar el box plot generado
5		Estimar histograma	El sistema deberá permitir estimar histogramas por atributos en set de datos dado
6		Visualizar histograma	El sistema deberá permitir visualizar el histograma generado

7		Estimar bar charts	El sistema deberá permitir estimar bar charts para atributos con distribución discreta
8		Visualizar bar chart	El sistema deberá permitir visualizar el bar chart generado
9		Estimar pie charts	El sistema deberá permitir estimar pie charts para atributos con distribución discreta

10		Visualizar pie charts	El sistema deberá permitir visualizar el pie charts generado
11		Estimar matrices de correlación	El sistema deberá permitir estimar matrices de correlación para un set de datos de interés
12		Visualizar heat map	El sistema deberá permitir visualizar el heat map generado relacionado a la matriz de correlación

13			Estimar scatter plot	El sistema deberá permitir estimar scatter plot para un set de datos de interés
14			Visualizar scatter plot	El sistema deberá permitir visualizar el scatter plot generado

Tabla 2.5: Funciones asociadas al módulo estadístico.

2.1.5. Funciones asociadas al set de datos

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción

1	ESD	Evaluación set de datos	Revisar correcto estado set de datos	El sistema deberá permitir revisar el correcto estado del set de datos
2			Revisar variables discretas en set de datos	El sistema deberá permitir revisar si el set de datos presenta variables discretas en sus atributos

3			Revisar formato de set de datos	El sistema deberá permitir revisar el formato del set de datos y chequear si viene según las características impuestas
4			Alojar set de datos en área de trabajo	El sistema deberá permitir alojar el set de datos en el área de trabajo del usuario

Tabla 2.6: Funciones asociadas al set de datos.

2.1.6. Funciones asociadas al módulo de algoritmos de aprendizaje no supervisado

<i>Funciones del Sistema</i>				
<i>#</i>	<i>ID Categoría</i>	<i>Categoría</i>	<i>Función</i>	<i>Descripción</i>
1	MANS	Modelos de Aprendizaje No Supervisado	Implementar algoritmo K-Means	El sistema deberá permitir implementar el algoritmo K-Means para entorno de aprendizaje no supervisado
2			Implementar algoritmo Mean Shift	El sistema deberá permitir implementar el algoritmo Mean Shift para entorno de aprendizaje no supervisado

3			Implementar algoritmo	El sistema deberá permitir implementar el algoritmo
			Affinity Propagation	Affinity Propagation para entorno de aprendizaje no supervisado
4			Implementar algoritmo	El sistema deberá permitir implementar el algoritmo
			DBScan	DBScan para entorno de aprendizaje no supervisado

5			Implementar algoritmo Aglomerativos	El sistema deberá permitir implementar el algoritmo Aglomerativos para entorno de aprendizaje no supervisado
6			Implementar algoritmo jerarquizado	El sistema deberá permitir implementar el algoritmo Jerarquizado para entorno de aprendizaje no supervisado

7			Implementar algoritmo SOM	El sistema deberá permitir implementar el algoritmo Self Organization Map para entorno de aprendizaje no supervisado
8			Evaluar las particiones generadas	El sistema deberá permitir la evaluación de las particiones generadas mediante Coeficientes determinados

9			Reportar resultados	El sistema deberá permitir reportar los resultados obtenidos para su posterior análisis
---	--	--	---------------------	-----------------------------------------------------------------------------------------

Tabla 2.7: Funciones asociadas a los módulos de clustering

2.1.7. Funciones asociadas al módulo de clasificaciones

<i>Funciones del Sistema</i>				
#	<i>ID Categoría</i>	<i>Categoría</i>	<i>Función</i>	<i>Descripción</i>

1			Implementar algoritmo Naive Bayes	El sistema deberá permitir implementar el algoritmo Naive Bayes para entorno de aprendizaje supervisado
2			Implementar algoritmo KNN	El sistema deberá permitir implementar el algoritmo KNN para entorno de aprendizaje supervisado
	MAS	<i>Modelos de Aprendizaje</i>		

Supervisado

3			Implementar algoritmo Random Forest	El sistema deberá permitir implementar el algoritmo Random Forest para entorno de aprendizaje supervisado
4			Implementar algoritmo AdaBoost	El sistema deberá permitir implementar el algoritmo AdaBoost para entorno de aprendizaje supervisado

5			Implementar algoritmo SVM	El sistema deberá permitir implementar el algoritmo SVM para entorno de aprendizaje supervisado
6			Implementar algoritmo NuSVC	El sistema deberá permitir implementar el algoritmo NuSVC para entorno de aprendizaje supervisado

7			Implementar algoritmo MLP	El sistema deberá permitir implementar el algoritmo MLP para entorno de aprendizaje supervisado
8			Implementar algoritmo Árboles de Decisión	El sistema deberá permitir implementar el algoritmo Árboles de decisión para entorno de aprendizaje supervisado

9		Implementar validación cruzada	El sistema deberá permitir implementar los algoritmos de validación cruzada para cada algoritmo
10		Implementar validación LOU	El sistema deberá permitir implementar los algoritmos de validación LOU para cada algoritmo
11		Estimar Curva ROC	El sistema deberá permitir estimar la curva roc para el modelo resultante

12	Estimar curva de validación	El sistema deberá permitir estimar la curva de validación para el modelo resultante
13	Estimar matriz de confusión	El sistema deberá permitir estimar la matriz de confusión para el modelo resultante

14			Estimar las medidas de desempeño	El sistema deberá permitir estimar las medidas de desempeño obtenidas para el modelo resultante
15			Alojar resultados en job	El sistema deberá permitir alojar los resultados obtenidos en el job relacionado al área de trabajo

Tabla 2.8: Funciones asociadas al módulo de clasificación.

2.1.8. Funciones asociadas al módulo de análisis de características

<i>Funciones del Sistema</i>				
<i>#</i>	<i>ID Categoría</i>	<i>Categoría</i>	<i>Función</i>	<i>Descripción</i>
1	AF	Análisis de	Implementar PCA	El sistema deberá permitir la implementación del algoritmo PCA y exponer los resultados generados
2		Características	Implementar Deformaciones de Espacio	El sistema deberá permitir la implementación de la evaluación de características mediante Importancia en Random Forest

3		Implementar algoritmos de Mutual Information	El sistema deberá permitir la implementación de los algoritmos referentes a mutual information
4		Implementar algoritmode correlaciones	El sistema deberá permitir implementar y analizar los resultados de análisis de correlaciones

5			Implementar algoritmos basados en distribuciones bayesianas	El sistema deberá permitir la implementación de métodos bayesianos para análisis de características
---	--	--	----------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------

Tabla 2.9: Funciones asociadas al módulo de análisis de características

2.1.9. Funciones asociada al Sistema de Colas

<i>Funciones del Sistema</i>				
<i>#</i>	<i>ID Categoría</i>	<i>Categoría</i>	<i>Función</i>	<i>Descripción</i>
1			Registrar procesos en sistema de colas	El sistema deberá permitir el registro de nuevos procesos en sistema de colas y establecer las configuraciones sobre el mismo

2	
3	

Editar proceso en sistema de cola	El sistema deberá permitir modificar los procesos encolados, cambiarles estado, modificar prioridades, etc
Eliminar proceso de sistema de cola	El sistema deberá permitir remover procesos en sistemas de colas e identificar las consecuencias del hecho.

4		Mostrar estado de cola	El sistema deberá permitir consultar los procesos en sistema de colas y exponer las características de ellos.
5		Notificar estados de procesos	El sistema deberá permitir notificar a los usuarios los estados de los procesos
6		Notificar finalización de trabajos	El sistema deberá permitir la notificación a los usuarios cuando un job finalice

7			Revisar procesos en cola	El sistema deberá permitir la revisión constante de los procesos encolados
---	--	--	---------------------------------	------------------------------------------------------------------------------------------

Tabla 2.10: Funciones asociadas al Sistema de Colas

2.1.10. Funciones asociadas al Sistema de Notificación

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción
1	SN	Sistema de notificaciones	Enviar mensajes de notificaciones a usuarios	El sistema deberá permitir la notificación de nuevos mensajes a usuarios vía email o notificación push

2			Configurar mensajes de notificaciones	El sistema deberá permitir configurar y condicionar las notificaciones que se emiten
3			Generar notificaciones globales	El sistema deberá permitir generar notificaciones generales a todos los usuarios existentes

Tabla 2.11: Funciones asociadas al Sistema de Notificaciones

2.1.11. Funciones asociadas al Sistema de Almacenamiento Persistente

<i>Funciones del Sistema</i>				
#	ID Categoría	Categoría	Función	Descripción

1	AP	Almacenamiento Persistente	Responder solicitudes de muestra de datos	El sistema deberá permitir responder a las solicitudes de consulta de datos desde el sistema de almacenamiento persistente
2			Registrar nuevos elementos	El sistema deberá permitir registrar nuevos elementos en el sistema de almacenamiento persistente

3			Modificar elementos en el sistema	El sistema deberá permitir editar registros y mantener la concordancia de las relaciones asociadas al sistema de almacenamiento persistente
4			Eliminar elementos en el sistema	El sistema deberá permitir eliminar registros en el sistema de almacenamiento persistente

Tabla 2.12: Funciones asociadas al Sistema de Almacenamiento Persistente

2.2. Atributos del Sistema

Los atributos del sistema, representan variables cuantitativas asociadas a características de éste, es decir, índices que permiten identificar que el software cumple con características deseables a modo general.

Cada uno de los posibles atributos que posee el sistema se exponen a continuación en la Tabla 2.13, donde se detalla el atributo con su ID y su descripción o forma de medición

Es importante asociar que cada una de las funciones expuestas cumple con a lo menos

<i>Atributos del Sistema</i>		
<i>ID</i>	<i>Item</i>	<i>Descripción medición</i>
D	Disponibilidad	99 % Up Time
P	Persistencia	99 % en base a configuración servidores
TE	Tiempo de Espera	Menor a 1 minuto por acción
U	Usabilidad	Menos de 5 clicks por acción
TF	Tolerancia a Fallos	1 Fallo cada 1000 ejecuciones
R	Respaldos	Data dispuesta en distintos servidores según configuración
CM	Conexiones Múltiples	Hasta 1000 conexiones múltiples
SM	Solicitudes Múltiples	Hasta 100 ejecuciones en sistema de colas
S	Seguridad	99.9 % tolerable a ingresos corrompidos

Tabla 2.13: Atributos presentes en el sistema

un atributo de los nombrados en la Tabla 2.13, la cruza de esta información, es decir los atributos por función se expone en la tabla 2.14

<i>Atributos por Función</i>			
#	ID	Función	Atributos
1	US	Registrar nuevo usuario	D, P, TE, U, TF
2		Editar usuario existente	D, P, TE, U, TF
3		Eliminar usuario existente	D, P, TE, U, TF
4		Visualizar usuario	D, P, TE, U, TF
5		Crear áreas de trabajo	D, P, TE, U, TF
6		Eliminar áreas de trabajo	D, P, TE, U, TF
7		Registrar roles	D, P, TE, U, TF
8		Editar roles existente	D, P, TE, U, TF
9		Eliminar rol existente	D, P, TE, U, TF
10		Visualizar roles	D, P, TE, U, TF

1	AC	Iniciar sesión de usuario	D, S, CM, TE
2		Autenticar usuario	D, S, CM, TE
3		Cerrar sesión	D, S, CM, TE
4		Recuperar cuenta de usuario	D, S, CM, TE, TF
5		Reestablecer cuenta de usuario	D, S, CM, TE, P
6		Modificar datos de acceso	D, S, CM, TE, TF
7		Notificación de cambios	D, S, CM, TE, TF
1	ES	Visualizar estadísticas de uso	U, SM, TE
2		Visualizar carga de servidores	U, SM, TE
3		Visualizar carga de sistema de colas	U, SM, TE
4		Visualizar estados de trabajos	U, SM, TE

5		Visualizar estadísticas de trabajos	U, SM, TE
1	AE	Estimar estadísticos en relación a la muestra	U, SM, TE, CM, TF
2		Visualizar estadísticos de datos	U, SM, TE, CM, TF
3		Estimar box plot	U, SM, TE, CM, TF
4		Visualizar box plot	U, SM, TE, CM, TF
5		Estimar histograma	U, SM, TE, CM, TF
6		Visualizar histograma	U, SM, TE, CM, TF
7		Estimar bar charts	U, SM, TE, CM, TF
8		Visualizar bar chart	U, SM, TE, CM, TF
9		Estimar pie charts	U, SM, TE, CM, TF
10		Visualizar pie charts	U, SM, TE, CM, TF

11		Estimar matrices de correlación	U, SM, TE, CM, TF
12		Visualizar heat map	U, SM, TE, CM, TF
13		Estimar scatter plot	U, SM, TE, CM, TF
14		Visualizar scatter plot	U, SM, TE, CM, TF
1	ESD	Revisar correcto estado set de datos	U, SM, TE, CM, TF
2		Revisar variables discretas en set de datos	U, SM, TE, CM, TF
3		Revisar formato de set de datos	U, SM, TE, CM, TF
4		Alojar set de datos en área de trabajo	U, SM, TE, CM, TF
1		Implementar algoritmo K-Means	U, SM, TE, CM, TF
2		Implementar algoritmo Mean Shift	U, SM, TE, CM, TF
3		Implementar algoritmo Affinity Propagation	U, SM, TE, CM, TF

4		Implementar algoritmo DBScan	U, SM, TE, CM, TF
5		Implementar algoritmo Aglomerativos	U, SM, TE, CM, TF
6		Implementar algoritmo jerarquizado	U, SM, TE, CM, TF
7		Implementar algoritmo SOM	U, SM, TE, CM, TF
8		Evaluar las particiones generadas	U, SM, TE, CM, TF
9		Reportar resultados	U, SM, TE, CM, TF
1		Implementar algoritmo Naive Bayes	U, SM, TE, CM, TF
2		Implementar algoritmo KNN	U, SM, TE, CM, TF
3		Implementar algoritmo Random Forest	U, SM, TE, CM, TF
4		Implementar algoritmo AdaBoost	U, SM, TE, CM, TF
5		Implementar algoritmo SVM	U, SM, TE, CM, TF

6	Implementar algoritmo NuSVC	U, SM, TE, CM, TF
7	Implementar algoritmo MLP	U, SM, TE, CM, TF
8	Implementar algoritmo Árboles de Decisión	U, SM, TE, CM, TF
9	Implementar validación cruzada	U, SM, TE, CM, TF
10	Implementar validación LOU	U, SM, TE, CM, TF
11	Estimar Curva ROC	U, SM, TE, CM, TF
12	Estimar curva de validación	U, SM, TE, CM, TF
13	Estimar matriz de confusión	U, SM, TE, CM, TF
14	Estimar las medidas de desempeño	U, SM, TE, CM, TF
15	Alojar resultados en job	U, SM, TE, CM, TF
1	Implementar PCA	U, SM, TE, CM, TF

2		Implementar Deformaciones de Espacio	U, SM, TE, CM, TF
3		Implementar algoritmos de Mutual Information	U, SM, TE, CM, TF
4		Implementar algoritmo de correlaciones	U, SM, TE, CM, TF
5		Implementar algoritmos basados en distribuciones bayesianas	U, SM, TE, CM, TF
1	SC	Registrar procesos en sistema de colas	TE, TF, SM, CM, S, R, P
2		Editar proceso en sistema de cola	TE, TF, SM, CM, S, R, P
3		Eliminar proceso de sistema de cola	TE, TF, SM, CM, S, R, P
4		Mostrar estado de cola	TE, TF, SM, CM, S, R, P
5		Notificar estados de procesos	TE, TF, SM, CM, S, R, P
6		Notificar finalización de trabajos	TE, TF, SM, CM, S, R, P

7		Revisar procesos en cola	TE, TF, SM, CM, S, R, P
1	SN	Enviar mensajes de notificaciones a usuarios	TF, CM, SM, S
2		Configurar mensajes de notificaciones	TF, CM, SM, S
3		Generar notificaciones globales	TF, CM, SM, S
1	AP	Responder solicitudes de muestra de datos	P, R, SM, CM, TF, TE
2		Registrar nuevos elementos	P, R, SM, CM, TF, TE
3		Modificar elementos en el sistema	P, R, SM, CM, TF, TE
4		Eliminar elementos en el sistema	P, R, SM, CM, TF, TE

Tabla 2.14: Tabla resumen de Atributos por Función

2.3. Actores y Usuarios

Los actores son aquellos entes que participan o interactúan con el software pero que no forman parte directa de éste, es decir, son agentes externos. En este diseño de software se han detectado los siguientes actores que se exponen en la Tabla 2.15.

Actores del Sistema		
Actor	ID	Descripción
Usuarios	U	Entidad externa que harán uso de los servicios y módulos que disponga el sistema
Almacenamiento Persistente	AP	Entidad que representa el entorno de almacenamiento persistente del sistema

Sistema de Colas	SC	Entidad externa que facilita el manejo de los procesos a ejecutarse de manera remota en forma de jobs
Scikit-learn	SCL	Entidad externa que facilita la implementación de los modelos de clasificación y clustering

Tabla 2.15: Actores que interactúan con el sistema

En la Tabla 2.15 se encuentran definidos los actores, además se aprecia que existe el actor Usuario. Sin embargo, se destaca que los usuarios pueden tener roles o perfiles, los cuales se exponen en la Tabla 2.16

2.4. Casos de Uso

Un caso de uso, representa la narrativa de un conjunto de acciones que suceden entre actores y el sistema, exponiendo los flujos de acciones o la lógica tras una acción, se asocian a funciones del sistema y tienen la finalidad de cumplir con alguna de éstas, normalmente se encuentran asociados a más casos de uso, generando un conjunto de referencias cruzadas, además se evalúa su funcionamiento mediante la determinación de

Perfiles del Sistema		
Rol	ID	Descripción
Administrador	A	Encargado de administrar las cuentas de usuario, los sistemas de acceso, estadísticas de uso, etc
Usuario Común	UC	Tiene acceso a los módulos y a la utilización de recursos que el sistema dispone
Root	R	Usuario con todos los privilegios de acceso

Tabla 2.16: Perfiles de usuario identificados

las pre y post condiciones.

A continuación se exponen todos los principales casos de uso para el sistema, los cuales se asocian la función a la cual pertenecen, además se agrupan según actor o acciones que engloban.

2.4.1. Casos de uso asociados a las acciones del Usuario administrador

Estos casos de uso se asocian a las acciones que tienen estrecha relación con las obligaciones que cumple un usuario con rol de administrador o root.

<i>Caso de Uso: CU01</i>	
Nombre	Registrar nuevo usuario
Actores	U, AP
Funciones Asociadas	US1, AP1, AP2
Objetivo	Insertar en el sistema de almacenamiento persistente un nuevo usuario
Pre Condiciones	Sistema de almacenamiento persistente sin registro de usuario

Post Condiciones	Usuario registrado en sistema, área de trabajo creada correctamente
Acciones del Actor	Acciones del Sistema
Administrador ingresa a formulario de nuevo registro de usuario	
	Sistema solicita los datos de usuario
Administrador completa y envía formulario	

	Sistema procesa y valida los datos
	Sistema registra nuevo dato en almacenamiento persistente (CU02)
	Sistema crea área de trabajo para nuevo usuario (CU03)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU02, CU03

Cursos alternativos	Usuario ya existe y el sistema lo notifica
	No se puede registrar el usuario por problemas en AP, se notifica

<i>Caso de Uso: CU06</i>	
Nombre	Editar usuario existente
Actores	U, AP
Funciones Asociadas	

Objetivo	Editar usuario existente en el sistema de almacenamiento persistente
Pre Condiciones	Sistema de almacenamiento persistente sin usuario modificado
Post Condiciones	Usuario registrado modificado de manera correcta
Acciones del Actor	Acciones del Sistema
Administrador ingresa a formulario de editar registro de usuario	

	Sistema solicita los datos de usuario
Administrador completa y envía formulario	
	Sistema procesa y valida los datos
	Sistema edita registro en almacenamiento persistente (CU07)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU07

Cursos alternativos	No se puede editar el usuario por problemas en AP, se notifica
----------------------------	-----------------------------------------------------------------------

<i>Caso de Uso: CU08</i>	
Nombre	Eliminar usuario existente
Actores	U, AP
Funciones Asociadas	
Objetivo	Eliminar usuario existente en el sistema de almacenamiento persistente

Pre Condiciones	Sistema de almacenamiento persistente con usuario registrado
Post Condiciones	Usuario registrado eliminado de manera correcta
Acciones del Actor	Acciones del Sistema
Administrador ingresa a formulario de eliminar registro de usuario	
	Sistema solicita los datos de usuario
Administrador completa y envía formulario	

	Sistema procesa y valida los datos
	Sistema elimina registro en almacenamiento persistente (CU09)
	Sistema elimina área de trabajo del usuario (CU10)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU09, CU10
Cursos alternativos	No se puede eliminar el usuario por problemas en AP, se notifica

<i>Caso de Uso: CU11</i>	
Nombre	Visualizar usuarios
Actores	U, AP
Funciones Asociadas	
Objetivo	Visualizar lista de usuarios o datos de usuario seleccionado
Pre Condiciones	Solicitud de visualización de usuarios
Post Condiciones	Solicitud respondida de visualización de usuarios

Acciones del Actor	Acciones del Sistema
Administrador ingresa a sección visualizar usuarios	
	Sistema recibe la solicitud
	Sistema consulta la data al Sistema de Almacenamiento persistente
	Sistema recupera la respuesta en formato JSON
	Sistema decodifica la información

	Sistema despliega la respuesta en la interfaz de usuario
Referencias Cruzadas	CU12
Cursos alternativos	No se puede mostrar la data, se notifica la acción

<i>Caso de Uso: CU13</i>	
Nombre	Registrar nuevo rol
Actores	U, AP
Funciones Asociadas	

Objetivo	Insertar en el sistema de almacenamiento persistente un nuevo rol
Pre Condiciones	Sistema de almacenamiento persistente sin registro de rol
Post Condiciones	Rol registrado en sistema
Acciones del Actor	Acciones del Sistema
Administrador ingresa a formulario de nuevo registro de rol	

	Sistema solicita los datos de rol
Administrador completa y envía formulario	
	Sistema procesa y valida los datos
	Sistema registra nuevo dato en almacenamiento persistente (CU02)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU02

Cursos alternativos	Rol ya existe y el sistema lo notifica
	No se puede registrar el rol por problemas en AP, se notifica

<i>Caso de Uso: CU14</i>	
Nombre	Editar rol existente
Actores	U, AP
Funciones Asociadas	
Objetivo	Editar rol existente en el sistema de almacenamiento persistente

Pre Condiciones	Sistema de almacenamiento persistente sin rol modificado
Post Condiciones	Rol editado modificado de manera correcta
Acciones del Actor	Acciones del Sistema
Administrador ingresa a formulario de editar registro de rol	
	Sistema solicita los datos de rol

Administrador completa y envía formulario	
	Sistema procesa y valida los datos
	Sistema edita registro en almacenamiento persistente (CU07)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU07
Cursos alternativos	No se puede editar el rol por problemas en AP, se notifica

<i>Caso de Uso: CU15</i>	
Nombre	Eliminar rol existente
Actores	U, AP
Funciones Asociadas	
Objetivo	Eliminar rol existente en el sistema de almacenamiento persistente
Pre Condiciones	Sistema de almacenamiento persistente con rol registrado

Post Condiciones	Rol registrado eliminado de manera correcta
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Administrador ingresa a formulario de eliminar registro de rol	
	Sistema solicita los datos de rol
Administrador completa y envía formulario	
	Sistema procesa y valida los datos

	Sistema elimina registro en almacenamiento persistente (CU09)
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	CU09
Cursos alternativos	No se puede eliminar el rol por problemas en AP, se notifica

<i>Caso de Uso: CU16</i>	
Nombre	Visualizar roles
Actores	U, AP

Funciones Asociadas	
Objetivo	Visualizar lista de roles registrados en el sistema de almacenamiento persistente
Pre Condiciones	Solicitud de visualización de roles
Post Condiciones	Solicitud respondida de visualización de roles
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Administrador ingresa a sección visualizar roles	

	Sistema recibe la solicitud
	Sistema consulta la data al Sistema de Almacenamiento persistente
	Sistema recupera la respuesta en formato JSON
	Sistema decodifica la información
	Sistema despliega la respuesta en la interfaz de usuario
Referencias Cruzadas	CU12

Cursos alternativos	No se puede mostrar la data, se notifica la acción
----------------------------	-----------------------------------------------------------

Tabla 2.17: Casos de uso asociados a los usuarios administrativo y root

2.4.2. Casos de uso asociados al uso de recursos

Estos casos de uso se asocian al uso de recursos en el servidor, espacios físicos de almacenamiento, áreas de trabajo para los usuarios, etc.

<i>Caso de Uso: CU03</i>	
Nombre	Crear área de trabajo de usuario registrado
Actores	U
Funciones Asociadas	
Objetivo	Crear área de trabajo a usuario registrado

Pre Condiciones	No existe área de trabajo ni habilitación de recursos
Post Condiciones	Se habilita recursos para usuario y se crea área de trabajo
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario solicita la creación de nueva área de trabajo	
	Sistema habilita el usuario y crea espacio de trabajo
	Sistema habilita los recursos para el nuevo usuario

	Sistema habilita al usuario en el proceso de colas
	Sistema habilita al usuario en el registro de correos
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	—
Cursos alternativos	—

Caso de Uso: CU10

Nombre	Eliminar área de trabajo de usuario registrado
Actores	U
Funciones Asociadas	
Objetivo	Eliminar área de trabajo a usuario registrado
Pre Condiciones	Si existe área de trabajo ni habilitación de recursos
Post Condiciones	Se elimina área de trabajo y deshabilitan los recursos asociados
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Usuario solicita la eliminación de área de trabajo	
	Sistema deshabilita el usuario y elimina espacio de trabajo
	Sistema elimina los recursos para el usuario
	Sistema deshabilita al usuario en el proceso de colas
	Sistema deshabilita al usuario en el registro de correos

	Sistema notifica el correcto término del proceso
Referencias Cruzadas	—
Cursos alternativos	—

<i>Caso de Uso: CU40</i>	
Nombre	Eliminar archivo en área de trabajo
Actores	U
Funciones Asociadas	

Objetivo	Eliminar archivo en área de trabajo
Pre Condiciones	Archivo existe en el área de trabajo
Post Condiciones	Archivo eliminado del área de trabajo
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario sube un archivo al área de trabajo	
	Sistema detecta fallas en el archivo

	Sistema elimina el archivo ingresado
	Sistema notifica el correcto término del proceso
Referencias Cruzadas	—
Cursos alternativos	—

<i>Caso de Uso: CU41</i>	
Nombre	Agregar archivo en área de trabajo
Actores	U

Funciones Asociadas	
Objetivo	Agregar archivo en área de trabajo
Pre Condiciones	Archivo existe en el área de trabajo parcial
Post Condiciones	Archivo agregado al área de trabajo
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario sube un archivo al área de trabajo	
	Sistema revisa el archivo

	Sistema determina correcto archivo ingresado
	Sistema aloja en el área de trabajo el set de datos revisado
Referencias Cruzadas	—
Cursos alternativos	—

Tabla 2.18: Casos de uso asociados al uso de recursos.

2.4.3. Casos de uso asociados al sistema de almacenamiento persistente

El sistema de almacenamiento persistente tiene la responsabilidad de controlar todas las acciones relacionadas con el respaldo de data y mantenimiento de información de configuración del sistema.

<i>Caso de Uso: CU02</i>	
Nombre	Insertar registro en Sistema de Almacenamiento persistente
Actores	AP
Funciones Asociadas	
Objetivo	Crear un nuevo registro en el sistema de almacenamiento persistente
Pre Condiciones	Sistema de almacenamiento persistente sin registro
Post Condiciones	Sistema de almacenamiento persistente con registro agregado
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

AP recibe una solicitud de registro de datos en formato JSON	
AP procesa conexión a la base de datos (CU04)	
	Sistema recibe los datos mediante proceso JSON
	Sistema decodifica los datos
AP inserta los datos en el entorno correspondiente	
	Sistema notifica el correcto término del proceso

AP establece desconexión del sistema BD (CU05)	
Referencias Cruzadas	CU04, CU05
Cursos alternativos	No se puede iniciar sesión en BD, se notifica y el proceso se encola

<i>Caso de Uso: CU04</i>	
Nombre	Iniciar sesión en base de datos
Actores	AP
Funciones Asociadas	

Objetivo	Iniciar nueva sesión en base de datos
Pre Condiciones	Sesión no iniciada en BD
Post Condiciones	Sesión creada e iniciada en BD
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP solicita iniciar sesión en BD	
	Sistema recibe solicitud
	Sistema procesa las credenciales
	Sistema establece conexión
	Sistema notifica inicio de sesión correcto
Referencias Cruzadas	—

Cursos alternativos	Credenciales de acceso incorrectas, se notifica mediante el sistema
	Sistema BD no disponible por múltiples conexiones activas, sistema notifica el evento

<i>Caso de Uso: CU05</i>	
Nombre	Cerrar sesión en BD
Actores	AP
Funciones Asociadas	

Objetivo	Cerrar sesión iniciada en base de datos
Pre Condiciones	Sesión en estado iniciada
Post Condiciones	Sesión en estado cerrada
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP solicita cerrar sesión	
	Sistema BD recibe solicitud
	Sistema cierra la sesión
	Sistema elimina los datos de memoria caché y registros activos

	Sistema notifica el correcto término del proceso
Referencias Cruzadas	—
Cursos alternativos	—

<i>Caso de Uso: CU07</i>	
Nombre	Editar registro en Sistema de Almacenamiento persistente
Actores	AP
Funciones Asociadas	

Objetivo	Editar un registro en el sistema de almacenamiento persistente
Pre Condiciones	Sistema de almacenamiento persistente sin registro editado
Post Condiciones	Sistema de almacenamiento persistente con registro editado
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP recibe una solicitud de edición de datos en formato JSON	
AP procesa conexión a la base de datos (CU04)	

	Sistema recibe los datos mediante proceso JSON
	Sistema decodifica los datos
AP edita los datos en el entorno correspondiente	
	Sistema notifica el correcto término del proceso
AP establece desconexión del sistema BD (CU05)	
Referencias Cruzadas	CU04, CU05

Cursos alternativos	No se puede iniciar sesión en BD, se notifica y el proceso se encola
----------------------------	-----------------------------------------------------------------------------

<i>Caso de Uso: CU09</i>	
Nombre	Eliminar registro en Sistema de Almacenamiento persistente
Actores	AP
Funciones Asociadas	
Objetivo	Eliminar un registro en el sistema de almacenamiento persistente

Pre Condiciones	<p>Sistema de almacenamiento persistente</p> <p>sin registro eliminado</p>
Post Condiciones	<p>Sistema de almacenamiento persistente</p> <p>con registro eliminado</p>
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
<p>AP recibe una solicitud de eliminación</p> <p>de datos en formato JSON</p>	
<p>AP procesa conexión a la base de</p> <p>datos (CU04)</p>	
	<p>Sistema recibe los datos mediante</p> <p>proceso JSON</p>

	Sistema decodifica los datos
AP elimina los datos en el entorno correspondiente	
	Sistema notifica el correcto término del proceso
AP establece desconexión del sistema BD (CU05)	
Referencias Cruzadas	CU04, CU05
Cursos alternativos	No se puede iniciar sesión en BD, se notifica y el proceso se encola

<i>Caso de Uso: CU12</i>	
Nombre	Mostrar registro en Sistema de Almacenamiento persistente
Actores	AP
Funciones Asociadas	
Objetivo	Mostrar un conjunto de registros en el sistema de almacenamiento persistente
Pre Condiciones	Solicitud de búsqueda de información en estado espera

Post Condiciones	Solicitud de búsqueda de información en estado finalizada
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP recibe la solicitud de búsqueda de datos	
AP procesa conexión a la base de datos (CU04)	
	Sistema recibe los datos mediante proceso JSON
	Sistema responde la solicitud con el correspondiente resultado obtenido

AP establece desconexión del sistema BD (CU05)	
Referencias Cruzadas	CU04, CU05
Cursos alternativos	No se puede iniciar sesión en BD, se notifica y el proceso se encola

Tabla 2.19: Casos de uso relacionados al sistema de almacenamiento persistente.

2.4.4. Casos de uso asociados al sistema de notificación

El sistema de notificación tiene la responsabilidad de establecer los protocolos de comunicación hacia los usuarios en caso de que se genere un evento asociado.

<i>Caso de Uso: CU21</i>	
Nombre	Notificar mensaje vía correo electrónico
Actores	AP, U
Funciones Asociadas	

Objetivo	Enviar correo electrónico a usuario con mensaje asociado
Pre Condiciones	Sistema de mensajería disponible
Post Condiciones	Sistema de mensajería con solicitud realizada
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP genera solicitud de envío de data	
	Sistema recibe la solicitud
	Sistema revisa la data
	Sistema crea el cuerpo del correo

	Sistema envía vía SMTP el correo al destino asociado
	Sistema registra en su log el proceso generado
	Sistema notifica el proceso finalizado de manera correcta
Referencias Cruzadas	—
Cursos alternativos	Sistema de mensajería no disponible, se notifica la acción
	Servidor de correos SMTP no habilitado, se notifica la acción

Tabla 2.20: Casos de uso relacionados al sistema de notificaciones

2.4.5. Casos de uso asociados al manejo de sesiones

El manejo de sesiones en cualquier sistema de información es una de las tareas más relevantes, debido a que facilita la autenticación de usuario y permite una restricción de las acciones que los usuarios puedan realizar en el software, además permite mantener un control de los usuarios y desarrollar estadísticas de uso.

<i>Caso de Uso: CU17</i>	
Nombre	Iniciar sesión
Actores	U, AP
Funciones Asociadas	
Objetivo	Ingresar al sistema de información mediante la validación de las credenciales de usuario
Pre Condiciones	Solicitud de inicio de sesión
Post Condiciones	Inicio de sesión creado

<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede al formulario login	
Usuario completa formulario	
	Sistema recupera la data del formulario
	Sistema consulta la data al sistema de almacenamiento persistente
	Sistema verifica la respuesta
	Sistema inicia la sesión y entrega token de acceso
Usuario validado correctamente ingresa al sistema	

Referencias Cruzadas	CU12
Cursos alternativos	No se puede acceder al AP, se notifica al usuario
	Credenciales incorrectas, se notifica al usuario

<i>Caso de Uso: CU18</i>	
Nombre	Autenticar usuario
Actores	U, AP
Funciones Asociadas	
Objetivo	Autenticar usuario en sistema

Pre Condiciones	Usuario registrado en sistema pero sin acceso
Post Condiciones	Usuario registrado en sistema y habilitado para acceder
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Administrador registra un nuevo usuario	
	Sistema responde solicitud
	Sistema inicia proceso de autenticación
	Sistema genera token de registro
	Sistema almacena registro en BD

	Sistema codifica credenciales de usuario
	Sistema habilita las credenciales para usuario registrado
Referencias Cruzadas	CU12
Cursos alternativos	No se puede acceder al AP, se notifica al usuario
	Credenciales incorrectas, se notifica al usuario

<i>Caso de Uso: CU19</i>	
Nombre	Cerrar sesión

Actores	U, AP
Funciones Asociadas	
Objetivo	<p>Salir del sistema de información</p> <p>mediante la validación de las</p> <p>credenciales de usuario</p>
Pre Condiciones	Solicitud de cierre de sesión
Post Condiciones	<p>Cierre de sesión procesado de</p> <p>manera correcta</p>
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario solicita cerrar sesión	
	Sistema recibe la solicitud

	Sistema registra la hora de acceso
	Sistema almacena en sus log el registro de finalizar sesión
	Sistema borra la data caché
	Sistema elimina token generado
	Sistema cierra sesión redireccionando a la interfaz principal
Referencias Cruzadas	CU12
Cursos alternativos	No se puede acceder al AP, se notifica al usuario

	Credenciales incorrectas, se notifica al usuario
--	---------------------------------------------------------

<i>Caso de Uso: CU20</i>	
Nombre	Recuperar cuenta de usuario
Actores	U, AP
Funciones Asociadas	
Objetivo	Recuperar acceso al sistema
Pre Condiciones	Solicitud realizada de recuperación de cuenta

Post Condiciones	Solicitud de recuperación de cuentas finalizada
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede al formulario de recuperación de cuenta	
Usuario completa formulario de recuperación	
	Sistema recibe la solicitud
	Sistema valida que la data ingresada sea correcta

	Sistema notifica vía correo electrónico link de recuperación
	Sistema habilita link de recuperación por un tiempo máximo de 30 minutos
Usuario recibe correo electrónico y accede a link	
Usuario completa formulario	
	Sistema recibe la solicitud
	Sistema actualiza la data
	Sistema autentica al usuario

	<p>Sistema notifica vía correo electrónico</p> <p>la recuperación de cuenta</p>
Referencias Cruzadas	CU12, CU21
Cursos alternativos	<p>No se puede acceder al AP,</p> <p>se notifica al usuario</p>
	<p>Credenciales incorrectas,</p> <p>se notifica al usuario</p>
	<p>Sistema de correos no se encuentra</p> <p>disponible, se notifica al usuario la acción</p>

Caso de Uso: CU22

Nombre	Reestablecer cuenta de usuario
Actores	U, AP
Funciones Asociadas	
Objetivo	Reestablecer cuenta de usuario
Pre Condiciones	Solicitud realizada de reestablecer cuenta de usuario
Post Condiciones	Solicitud de reestablecer cuenta finalizada
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede al formulario de recuperación de cuenta	

Usuario completa formulario de recuperación	
	Sistema recibe la solicitud
	Sistema valida que la data ingresada sea correcta
	Sistema notifica vía correo electrónico link de recuperación
	Sistema habilita link de recuperación por un tiempo máximo de 30 minutos
Usuario recibe correo electrónico y accede a link	
Usuario completa formulario	

	Sistema recibe la solicitud
	Sistema actualiza la data
	Sistema autentica al usuario
	Sistema notifica vía correo electrónico la recuperación de cuenta
Referencias Cruzadas	CU12, CU21
Cursos alternativos	No se puede acceder al AP, se notifica al usuario
	Credenciales incorrectas, se notifica al usuario
	Sistema de correos no se encuentra disponible, se notifica al usuario la acción

<i>Caso de Uso: CU23</i>	
Nombre	Modificar datos de acceso
Actores	U, AP
Funciones Asociadas	
Objetivo	<p>Modificar los valores de las credenciales</p> <p>para un usuario</p>
Pre Condiciones	Credenciales de usuario no modificadas
Post Condiciones	Credenciales de usuario modificadas
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
<p>Usuario accede al formulario de</p> <p>cambio de contraseña</p>	

Usuario completa formulario de cambio	
	Sistema recibe la solicitud
	Sistema valida que la data ingresada sea correcta
	Sistema notifica vía correo electrónico el cambio generado
	Sistema actualiza la data
	Sistema autentica al usuario
	Sistema notifica vía correo electrónico la recuperación de cuenta
Referencias Cruzadas	CU12, CU21

Cursos alternativos	No se puede acceder al AP, se notifica al usuario
	Credenciales incorrectas, se notifica al usuario
	Sistema de correos no se encuentra disponible, se notifica al usuario la acción

Tabla 2.21: Casos de uso asociados al manejo de sesiones en el sistema

2.4.6. Casos de uso asociados al módulo estadísticas de uso

Estos casos de uso están relacionados a las estadísticas que implican el uso del software por parte de los usuarios y la carga del servidor.

<i>Caso de Uso: CU24</i>	
Nombre	Visualizar estadísticas de uso
Actores	AP, U
Funciones Asociadas	

Objetivo	Visualizar estadísticas de uso del servicio por usuario
Pre Condiciones	Solicitud de estadísticas de uso realizada
Post Condiciones	Solicitud de estadísticas de uso respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a sección de estadísticas de uso	
	Sistema recibe la solicitud

	Sistema consulta a AP para solicitud de información
AP recibe la solicitud y la procesa	
	Sistema recibe respuesta AP
	Sistema procesa respuesta
	Sistema decodifica respuesta en formato JSON
	Sistema genera tabla resumen de estadísticas recibidas
	Sistema genera gráfico de reportes con el uso

Referencias Cruzadas	CU12, CU25, CU26
Cursos alternativos	—

<i>Caso de Uso: CU27</i>	
Nombre	Visualizar carga de servidores
Actores	AP, U
Funciones Asociadas	
Objetivo	Visualizar carga de servidores en tiempo real
Pre Condiciones	Solicitud de estado de servidores

Post Condiciones	Solicitud de estado de servidores respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a sección de carga de servidores	
	Sistema recibe la solicitud
	Sistema consulta a AP para solicitud de información
AP recibe la solicitud y la procesa	
	Sistema recibe respuesta AP
	Sistema procesa respuesta

	Sistema decodifica respuesta en formato JSON
	Sistema genera tabla resumen de estadísticas recibidas
	Sistema genera gráfico de reportes con el uso
Referencias Cruzadas	CU12, CU25, CU26
Cursos alternativos	—

Caso de Uso: CU28

Nombre	Visualizar carga de sistema de colas
Actores	AP, U
Funciones Asociadas	
Objetivo	Visualizar carga del sistema de colas de procesos
Pre Condiciones	Solicitud de carga del sistema de colas generada
Post Condiciones	Solicitud de carga del sistema de colas respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Usuario accede a sección de carga de sistema de colas	
	Sistema recibe la solicitud
	Sistema consulta a AP para solicitud de información
AP recibe la solicitud y la procesa	
	Sistema recibe respuesta AP
	Sistema procesa respuesta
	Sistema decodifica respuesta en formato JSON

	Sistema genera tabla resumen de estadísticas recibidas
	Sistema genera gráfico de reportes con el uso
Referencias Cruzadas	CU12, CU25, CU26
Cursos alternativos	—

<i>Caso de Uso: CU29</i>	
Nombre	Visualizar estados de trabajos
Actores	AP, U
Funciones Asociadas	

Objetivo	Visualizar estados de trabajos
Pre Condiciones	Solicitud de estados de trabajos en servidor generada
Post Condiciones	Solicitud de estados de trabajos en servidor respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a sección de estados de trabajos	
	Sistema recibe la solicitud
	Sistema consulta a AP para solicitud de información

AP recibe la solicitud y la procesa	
	Sistema recibe respuesta AP
	Sistema procesa respuesta
	Sistema decodifica respuesta en formato JSON
	Sistema genera tabla resumen de estadísticas recibidas
	Sistema genera gráfico de reportes con el uso
Referencias Cruzadas	CU12, CU25, CU26
Cursos alternativos	—

<i>Caso de Uso: CU30</i>	
Nombre	Visualizar estadísticas de trabajos
Actores	AP, U
Funciones Asociadas	
Objetivo	Visualizar estadísticas de trabajos
Pre Condiciones	Solicitud de estadísticas de trabajos en servidor generada
Post Condiciones	Solicitud de estadísticas de trabajos en servidor respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Usuario accede a sección de estadísticas de trabajos	
	Sistema recibe la solicitud
	Sistema consulta a AP para solicitud de información
AP recibe la solicitud y la procesa	
	Sistema recibe respuesta AP
	Sistema procesa respuesta
	Sistema decodifica respuesta en formato JSON

	Sistema genera tabla resumen de estadísticas recibidas
	Sistema genera gráfico de reportes con el uso
Referencias Cruzadas	CU12, CU25, CU26
Cursos alternativos	—

Tabla 2.22: Casos de uso asociados al módulo estadístico

2.4.7. Casos de uso asociados a la generación de reportes

Estos casos de uso se encuentran relacionados al manejo de reportes y la generación de tablas resúmenes y gráficos de interés.

<i>Caso de Uso: CU25</i>	
Nombre	Generar tabla resumen de estadísticos
Actores	AP

Funciones Asociadas	
Objetivo	Generar tabla resumen responsive y con opciones de descarga
Pre Condiciones	Tabla resumen sin datos disponibles
Post Condiciones	Tabla resumen generada correcta
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP responde la solicitud de estadísticas resumen	
	Sistema recibe la respuesta y procesa el formato JSON

	<p>Sistema carga la data en estructura</p> <p>dataTable</p>
	<p>Sistema habilita los botones de acción</p> <p>en el datatable</p>
	<p>Sistema carga la tabla en el div</p> <p>correspondiente</p>
Referencias Cruzadas	—
Cursos alternativos	—

<i>Caso de Uso: CU26</i>	
Nombre	Generar gráfico resumen de estadísticos

Actores	AP
Funciones Asociadas	
Objetivo	Generar gráfico resumen responsive y con opciones de descarga
Pre Condiciones	Gráfico resumen sin datos disponibles
Post Condiciones	Gráfico resumen generado correctamente
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
AP responde la solicitud de estadísticas resumen	
	Sistema recibe la respuesta y procesa el formato JSON

	Sistema carga la data en estructura JSON
	Sistema habilita los botones de acción en el Higcharts plugin
	Sistema carga el gráfico en el div correspondiente
Referencias Cruzadas	—
Cursos alternativos	—

Tabla 2.23: Casos de uso asociados a reportes

2.4.8. Casos de uso relacionados al módulo de análisis estadístico

Estos casos de uso cumplen con los requerimientos asociados al análisis estadístico de los set de datos, estos contemplan desarrollo de histogramas, matrices de heat map, estadísticos resúmenes, box plot, entre los principales.

<i>Caso de Uso: CU31</i>	
Nombre	Estimar estadísticos en relación a la muestra
Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar estadísticos en relación a la muestra
Pre Condiciones	Solicitud de estadísticos para set de datos de interés generada
Post Condiciones	Solicitud de estadísticos para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Usuario accede a la sección estadísticas en set de datos	
Usuario solicita las estadísticas resúmenes	
	Sistema recibe solicitud
	Sistema procesa la matriz de información
	Sistema estima promedio, varianza, desviación, máximos y mínimos en la muestra
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera tabla resumen

	Sistema notifica proceso generado de manera correcta
	Sistema despliega cuadro resumen con los estadísticos generados
Referencias Cruzadas	CU12, CU25
Cursos alternativos	—

<i>Caso de Uso: CU32</i>	
Nombre	Estimar box plot
Actores	AP, U
Funciones Asociadas	

Objetivo	Estimar box plot en relación a la muestra
Pre Condiciones	Solicitud de box plot para set de datos de interés generada
Post Condiciones	Solicitud de box plot para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	
Usuario solicita el box plot de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información

	Sistema estima el box plot para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta
	Sistema despliega gráfico resumen con el box plot generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

<i>Caso de Uso: CU33</i>	
Nombre	Estimar Histograma
Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar histogramas en relación a la muestra
Pre Condiciones	Solicitud de histogramas para set de datos de interés generada
Post Condiciones	Solicitud de histogramas para set de datos de interés respondida

<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	
Usuario solicita histogramas de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información
	Sistema estima los histogramas para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON

	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta
	Sistema despliega gráfico resumen con el histograma generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

<i>Caso de Uso: CU34</i>	
Nombre	Estimar Barcharts

Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar barcharts en relación a la muestra
Pre Condiciones	Solicitud de barcharts para set de datos de interés generada
Post Condiciones	Solicitud de barcharts para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	

Usuario solicita barcharts de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información
	Sistema estima los barcharts para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta

	Sistema despliega gráfico resumen con el barcharts generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

<i>Caso de Uso: CU35</i>	
Nombre	Estimar Piecharts
Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar piecharts en relación a la muestra

Pre Condiciones	Solicitud de piecharts para set de datos de interés generada
Post Condiciones	Solicitud de piecharts para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	
Usuario solicita piecharts de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información

	Sistema estima los piecharts para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta
	Sistema despliega gráfico resumen con el piecharts generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

<i>Caso de Uso: CU36</i>	
Nombre	Estimar Matriz de correlación de los atributos
Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar matriz de correlación en relación a la muestra
Pre Condiciones	Solicitud de matriz de correlación para set de datos de interés generada
Post Condiciones	Solicitud de matriz de correlación para set de datos de interés respondida

<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	
Usuario solicita matriz de correlación de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información
	Sistema estima los matriz de correlación para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON

	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta
	Sistema despliega tabla resumen con el matriz de correlación generado
Referencias Cruzadas	CU12, CU25
Cursos alternativos	—

Caso de Uso: CU37

Nombre	Estimar Matriz de correlación de los atributos en forma de heat map
Actores	AP, U
Funciones Asociadas	
Objetivo	Estimar matriz de correlación en relación a la muestra visualizando heat map
Pre Condiciones	Solicitud de heat map para set de datos de interés generada
Post Condiciones	Solicitud de heat map para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Usuario accede a la sección estadísticas en set de datos	
Usuario solicita visualizar matriz de correlación de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información
	Sistema estima los matriz de correlación para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera gráfico resumen

	Sistema notifica proceso generado de manera correcta
	Sistema despliega tabla resumen con el matriz de correlación generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

<i>Caso de Uso: CU38</i>	
Nombre	Estimar scatter plot de los atributos
Actores	AP, U
Funciones Asociadas	

Objetivo	Estimar scatter plot de los atributos
Pre Condiciones	Solicitud de scatter plot para set de datos de interés generada
Post Condiciones	Solicitud de scatter plot para set de datos de interés respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a la sección estadísticas en set de datos	
Usuario solicita scatter plot de la muestra	
	Sistema recibe solicitud
	Sistema procesa la matriz de información

	Sistema estima los scatter plot para todos los atributos en el set de datos
	Sistema genera respuesta en formato JSON
	Sistema recibe respuesta y genera gráfico resumen
	Sistema notifica proceso generado de manera correcta
	Sistema despliega tabla resumen con el matriz de correlación generado
Referencias Cruzadas	CU12, CU26
Cursos alternativos	—

Tabla 2.24: Casos de uso relacionados al módulo de análisis estadístico

2.4.9. Casos de uso asociados al módulo de revisión de los datos de entrada

Estos casos de uso están estrechamente relacionados a la validación de los set de datos de entrada que ingrese el usuario.

<i>Caso de Uso: CU39</i>	
Nombre	Revisar correcto estado set de datos
Actores	AP, U
Funciones Asociadas	
Objetivo	Revisar correcto estado set de datos
Pre Condiciones	Set de datos no revisado
Post Condiciones	Set de datos revisado y solicitud respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario sube un set de datos a su área de trabajo	
	Sistema recibe set de datos

	Sistema inicia revisión del set de datos
	Sistema revisa elementos nulos
	Sistema revisa elementos incorrectos
	Sistema revisa disconformidades en la matriz entregada
	Sistema revisa existencia de atributos con variable discreta
	Sistema notifica el resultado de la revisión
	Sistema genera resumen del proceso

	Sistema aloja set de datos en área de trabajo de usuario
Referencias Cruzadas	CU41
Cursos alternativos	Si set de datos no se acepta, se procede a eliminar el archivo, CU40

Tabla 2.25: Casos de uso relacionados a la revisión del set de datos

2.4.10. Casos de uso asociados al módulo de clustering

Estos casos de uso cumplen con las funciones relacionadas a la aplicación de algoritmos de clustering en set de datos.

<i>Caso de Uso: CU42</i>	
Nombre	Implementar algoritmo K-Means
Actores	SCL, U

Funciones Asociadas	
Objetivo	Implementar algoritmo K-Means
Pre Condiciones	Solicitud de k-means generada
Post Condiciones	Solicitud de k-means respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con K-Means	
	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	

	Sistema recibe el formulario
	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	
	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla

	<p>Sistema notifica el correcto funcionamiento del algoritmo</p>
Referencias Cruzadas	—
Cursos alternativos	<p>No se puede aplicar algoritmo, se notifica por las vías establecidas</p>

<i>Caso de Uso: CU43</i>	
Nombre	Implementar algoritmo Mean Shift
Actores	SCL, U
Funciones Asociadas	
Objetivo	Implementar algoritmo Mean Shift

Pre Condiciones	Solicitud de Mean Shift generada
Post Condiciones	Solicitud de Mean Shift respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con Mean Shift	
	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario

	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	
	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla

	Sistema notifica el correcto funcionamiento del algoritmo
Referencias Cruzadas	—
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

<i>Caso de Uso: CU44</i>	
Nombre	Implementar algoritmo Affinity Propagation
Actores	SCL, U
Funciones Asociadas	

Objetivo	<p>Implementar algoritmo Affinity</p> <p>Propagation</p>
Pre Condiciones	<p>Solicitud de Affinity</p> <p>Propagation generada</p>
Post Condiciones	<p>Solicitud de Affinity</p> <p>Propagation respondida</p>
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
<p>Usuario solicita aplicar clustering con</p> <p>Affinity Propagation</p>	

	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario
	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	

	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla
	Sistema notifica el correcto funcionamiento del algoritmo
Referencias Cruzadas	—
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

<i>Caso de Uso: CU45</i>	
Nombre	Implementar algoritmo DBScan
Actores	SCL, U
Funciones Asociadas	
Objetivo	Implementar algoritmo DBScan
Pre Condiciones	Solicitud de DBScan generada
Post Condiciones	Solicitud de DBScan respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con DBScan	

	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario
	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	

	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla
	Sistema notifica el correcto funcionamiento del algoritmo
Referencias Cruzadas	—
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

<i>Caso de Uso: CU46</i>	
Nombre	Implementar algoritmo Aglomerativos
Actores	SCL, U
Funciones Asociadas	
Objetivo	Implementar algoritmo Aglomerativos
Pre Condiciones	Solicitud de Aglomerativos generada
Post Condiciones	Solicitud de Aglomerativos respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con Aglomerativos	

	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario
	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	

	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla
	Sistema notifica el correcto funcionamiento del algoritmo
Referencias Cruzadas	—
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

Caso de Uso: CU48

Nombre	Implementar algoritmo jerarquizado
Actores	SCL, U
Funciones Asociadas	
Objetivo	Implementar algoritmo jerarquizado
Pre Condiciones	Solicitud de jerarquizado generada
Post Condiciones	Solicitud de jerarquizado respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con jerarquizado	

	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario
	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	
	Sistema recibe respuestas y resultados generados

	Sistema procesa la data
	Sistema despliega el resultado en pantalla
	Sistema notifica el correcto funcionamiento del algoritmo
Referencias Cruzadas	—
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

<i>Caso de Uso: CU49</i>	
Nombre	Implementar algoritmo SOM
Actores	SCL, U

Funciones Asociadas	
Objetivo	Implementar algoritmo SOM
Pre Condiciones	Solicitud de SOM generada
Post Condiciones	Solicitud de SOM respondida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con SOM	
	Sistema recibe solicitud y despliega formulario con los parámetros necesarios
Usuario completa el formulario	
	Sistema recibe el formulario

	Sistema procesa la data y envía para aplicación de algoritmo
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL genera respuesta de resultados	
SCL genera evaluación de resultados	
	Sistema recibe respuestas y resultados generados
	Sistema procesa la data
	Sistema despliega el resultado en pantalla

	<p>Sistema notifica el correcto funcionamiento del algoritmo</p>
Referencias Cruzadas	—
Cursos alternativos	<p>No se puede aplicar algoritmo, se notifica por las vías establecidas</p>

<i>Caso de Uso: CU50</i>	
Nombre	Evaluar las particiones generadas
Actores	SCL, U
Funciones Asociadas	
Objetivo	Evaluar las particiones generadas

Pre Condiciones	Particiones generadas mediante Clustering sin evaluación
Post Condiciones	Particiones generadas mediante Clustering evaluadas de manera correcta
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario solicita aplicar clustering con algunos de los algoritmos	
	Sistema recibe la solicitud

	Sistema procesa la solicitud y la envía al SCL
SCL recibe solicitud y aplica algoritmo con parámetros seleccionados	
SCL evalúa las particiones mediante calinski-harabazz	
SCL evalúa las particiones mediante coeficiente de siluetas	
SCL evalúa las particiones mediante análisis estadísticos	
SCL reporta resultados	

	Sistema recibe los resultados
	Sistema despliega resultados en interfaz mediante uso de data table
	Sistema despliega datatable y expone los resultados
	Sistema notifica el correcto término de proceso
Referencias Cruzadas	CU25
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

<i>Caso de Uso: CU51</i>	
Nombre	Generar fase exploratoria de algoritmos de clustering
Actores	SC, U
Funciones Asociadas	
Objetivo	Generar fase exploratoria de algoritmos de clustering
Pre Condiciones	Solicitud de fase exploratoria en algoritmos de clustering generada
Post Condiciones	Solicitud de fase exploratoria en algoritmos de clustering procesada

<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Usuario accede a módulo de clustering	
Usuario accede a fase exploratoria	
	Sistema recibe solicitud
	Sistema prepara el Job
	Sistema lanza job
	Sistema notifica al usuario que el job ha sido lanzado
SC evalúa el estado del job	

SC finaliza job	
SC notifica finalización de job	
	Sistema recibe notificación
	Sistema prepara resumen de resultados y particiones generadas
	Sistema despliega datatable y expone los resultados
	Sistema notifica al usuario el resultado final
Referencias Cruzadas	CU52, CU53, CU25
Cursos alternativos	No se puede aplicar algoritmo, se notifica por las vías establecidas

Tabla 2.26: Casos de uso relacionados a los módulos de clustering

2.4.11. Casos de uso relacionados al lanzamiento de Jobs

Los jobs representan acciones en el sistema que serán ejecutadas en segundo plano y será notificado el término de cada procesos.

<i>Caso de Uso: CU52</i>	
Nombre	Crear un nuevo Job
Actores	SC
Funciones Asociadas	
Objetivo	Crear un nuevo Job
Pre Condiciones	Job no creado
Post Condiciones	Job creado y lanzado al sistema de colas
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

SC recibe petición de crear nuevo job	
SC recibe la data	Sistema procesa la información y crea el job
	Sistema encola el job
	Sistema almacena en sistema persistente los atributos del job
	Sistema notifica mediante correo electrónico el estado del job
	Sistema habilita item para evaluar avance del job
Referencias Cruzadas	CU25, CU02, CU21

Cursos alternativos	—
----------------------------	---

<i>Caso de Uso: CU53</i>	
Nombre	Finalizar un Job
Actores	SC
Funciones Asociadas	
Objetivo	Finalizar un Job
Pre Condiciones	Job en estado de ejecución o en espera
Post Condiciones	Job en estado finalizado
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

SC recibe petición de finalizar job	
SC recibe la data	Sistema procesa la información y finaliza el job
	Sistema des encola el job
	Sistema almacena en sistema persistente los atributos del job
	Sistema notifica mediante correo electrónico el estado del job
Referencias Cruzadas	CU25, CU02
Cursos alternativos	—

<i>Caso de Uso: CU75</i>	
Nombre	Editar un Job
Actores	SC
Funciones Asociadas	
Objetivo	Editar un Job
Pre Condiciones	Job en estado de ejecución o en espera
Post Condiciones	Job con cambio de estado generado
Acciones del Actor	Acciones del Sistema

SC recibe petición de editar estado de job	
SC recibe la data	Sistema procesa la información y edita el job
	Sistema cambia estado de proceso en gestor de colas
	Sistema almacena en sistema persistente los atributos del job
	Sistema notifica mediante correo electrónico el estado del job
Referencias Cruzadas	CU25, CU02

Cursos alternativos	—
----------------------------	---

<i>Caso de Uso: CU76</i>	
Nombre	Mostrar estado de cola
Actores	SC
Funciones Asociadas	
Objetivo	Mostrar estado de cola
Pre Condiciones	Mostrar el estado de los trabajos en el sistema de colas
Post Condiciones	Solicitud de visualización de estado de colas en espera

<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
SC recibe petición de mostrar estado de colas	
SC recupera el estado de los jobs	
SC procesa la información en formato JSON	
SC envía respuesta	
	Sistema recibe respuesta
	Sistema crea gráfico resumen con los procesos y sus estados

	Sistema crea datatable con el resumen de los procesos
	Sistema despliega los resultados en la vista
Referencias Cruzadas	CU25, CU02, CU26
Cursos alternativos	—

<i>Caso de Uso: CU77</i>	
Nombre	Notificar estados de procesos
Actores	SC
Funciones Asociadas	

Objetivo	Notificar estados de procesos
Pre Condiciones	Notificar vía correo electrónico el estado de un proceso
Post Condiciones	Solicitud de notificación de estado de Job
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
Sistema recibe solicitud de notificación de estado de proceso	
Sistema chequea el estado y lo registra	
Sistema envía respuesta en formato JSON	

	Sistema recibe respuesta
	Sistema envía petición de notificación al gestor de correos
	Sistema genera correo y notifica al usuario
	Sistema registra el estado del proceso en almacenamiento persistente
	Sistema registra el estado del proceso en sistema Log
Referencias Cruzadas	CU02, CU21
Cursos alternativos	—

<i>Caso de Uso: CU78</i>	
Nombre	Notificar finalización de Job
Actores	SC
Funciones Asociadas	
Objetivo	Notificar finalización de Job
Pre Condiciones	Notificar vía correo electrónico la finalización de un proceso
Post Condiciones	Solicitud de notificación de estado de Job
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>

Sistema recibe solicitud de notificación de finalización de proceso	
Sistema chequea el estado y lo registra	
Sistema envía respuesta en formato JSON	
SC finaliza job	
	Sistema recibe respuesta
	Sistema envía petición de notificación al gestor de correos
	Sistema genera correo y notifica al usuario
	Sistema registra el estado del proceso en almacenamiento persistente

	Sistema registra el estado del proceso en sistema Log
Referencias Cruzadas	CU02, CU21
Cursos alternativos	—

<i>Caso de Uso: CU79</i>	
Nombre	Revisar procesos en cola
Actores	SC
Funciones Asociadas	
Objetivo	Revisar procesos en cola

Pre Condiciones	Revisar procesos en cola y determinar si es necesario cambiar estados
Post Condiciones	Revisión de procesos requerida
<i>Acciones del Actor</i>	<i>Acciones del Sistema</i>
SC recibe solicitud de revisar sus procesos	
SC revisa los estados de todos los procesos	
SC evalúa el tiempo de espera de cada proceso	
SC evalúa el tiempo de ejecución de cada proceso	
SC evalúa la prioridad de cada proceso	
SC determina qué procesos finalizar	
SC finaliza jobs	

SC genera resumen de revisión	
	Sistema recibe resumen
	Sistema notifica los cambios de estado según corresponda
	Sistema almacena en sistema persistente los atributos del job
	Sistema registra el estado del proceso en sistema Log
Referencias Cruzadas	CU02, CU21
Cursos alternativos	—

Tabla 2.27: Casos de uso asociados al lanzamiento de Jobs

2.4.12. Diagramas de casos de uso

2.5. Diagramas de secuencia o colaboración

2.6. Conceptos

2.6.1. Modelo Conceptual

2.7. Entidades

2.7.1. Modelo de Entidades

3. Diseño

3.1. Arquitectura de Software

3.2. Diagramas de Interacción

3.3. Diagrama de Clases

3.4. Diagramas de Estado

4. Planificación

4.1. Etapas del Proyecto

Bibliografía

- [1] *Data Mining Curriculum*. ACM SIGKDD. 2006-04-30. Retrieved 2014-01-27.
- [2] Clifton, Christopher (2010). *Encyclopedia Britannica: Definition of Data Mining*. Retrieved 2010-12-09.
- [3] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Retrieved 2012-08-07.
- [4] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). *From Data Mining to Knowledge Discovery in Databases* (PDF). Retrieved 17 December 2008.