

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos

Laboratorio

1

Introducción a la Inteligencia de Negocios

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo principal el que el estudiante se familiarice con la herramienta KNIME.

II

TEMAS A TRATAR

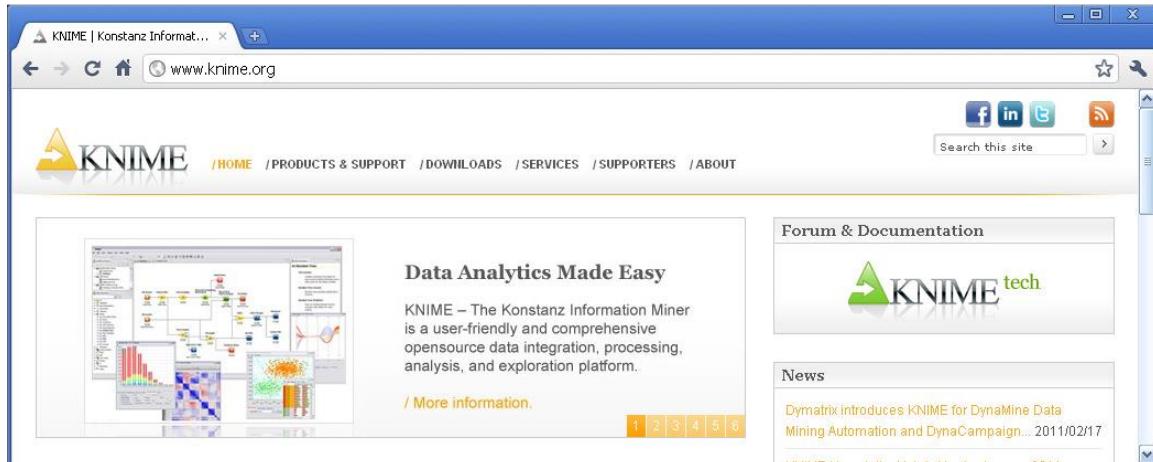
- Tema 1. Introducción a la Inteligencia de Negocios

III

MARCO TEÓRICO

- KNIME. Es una herramienta de minería de datos basada en software libre (<http://www.knime.org/>).
- Presenta algunas características interesantes para IN pues integra otras plataformas como BIRT (para creación de informes), WEKA (para minería de datos), R y Python (para análisis estadísticos y visualización) además de otras extensiones para lectura de datos, ETL, generación de informes, visualización e incluso Big Data.
- Aunque no se trate de una plataforma originalmente diseñada para IN, resulta muy adecuada como herramienta docente de iniciación por el tamaño comedido del software y su facilidad de instalación y configuración.

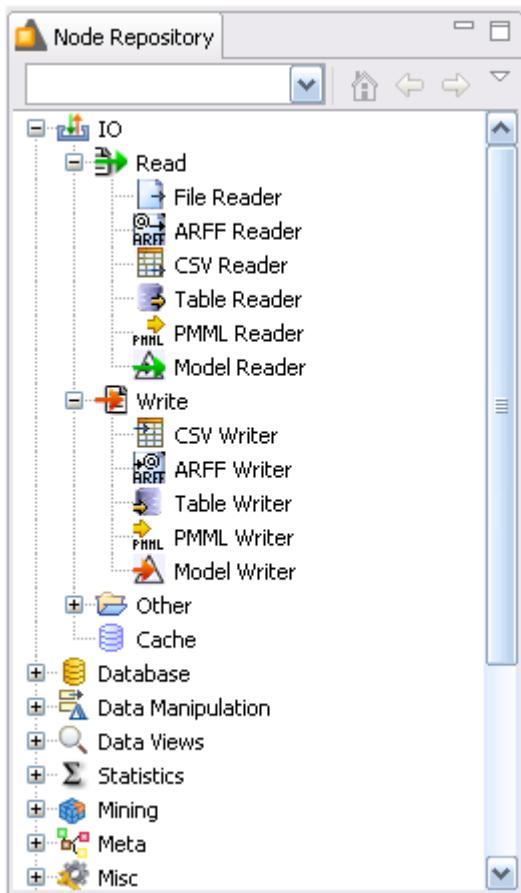
Introducción a KNIME
KoNstanz Information MinEr
<http://www.knime.org/>



KNIME es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos. KNIME fue desarrollado originalmente en el departamento de bioinformática y minería de datos de la Universidad de Constanza, Alemania, bajo la supervisión del profesor Michael Berthold. En la actualidad, la empresa KNIME.com GmbH, radicada en Zúrich, Suiza, continúa su desarrollo, además de prestar servicios de formación y consultoría.

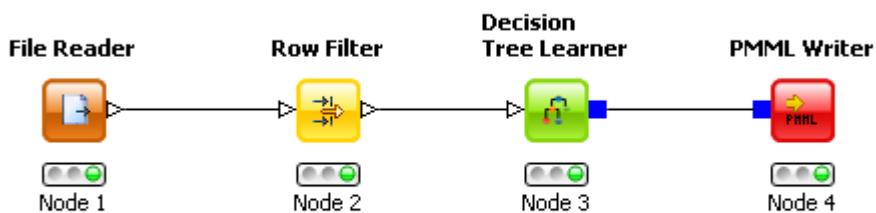
KNIME está desarrollado sobre la plataforma Eclipse y programado, esencialmente, en Java. Como otros entornos de este tipo, algunos de los cuales aparecen referenciados al final de este documento, su uso se basa en el diseño de un flujo de ejecución que plasme las distintas etapas de un proyecto de minería de datos.

Para ello, KNIME proporciona distintos nodos agrupados en fichas, como, por ejemplo:



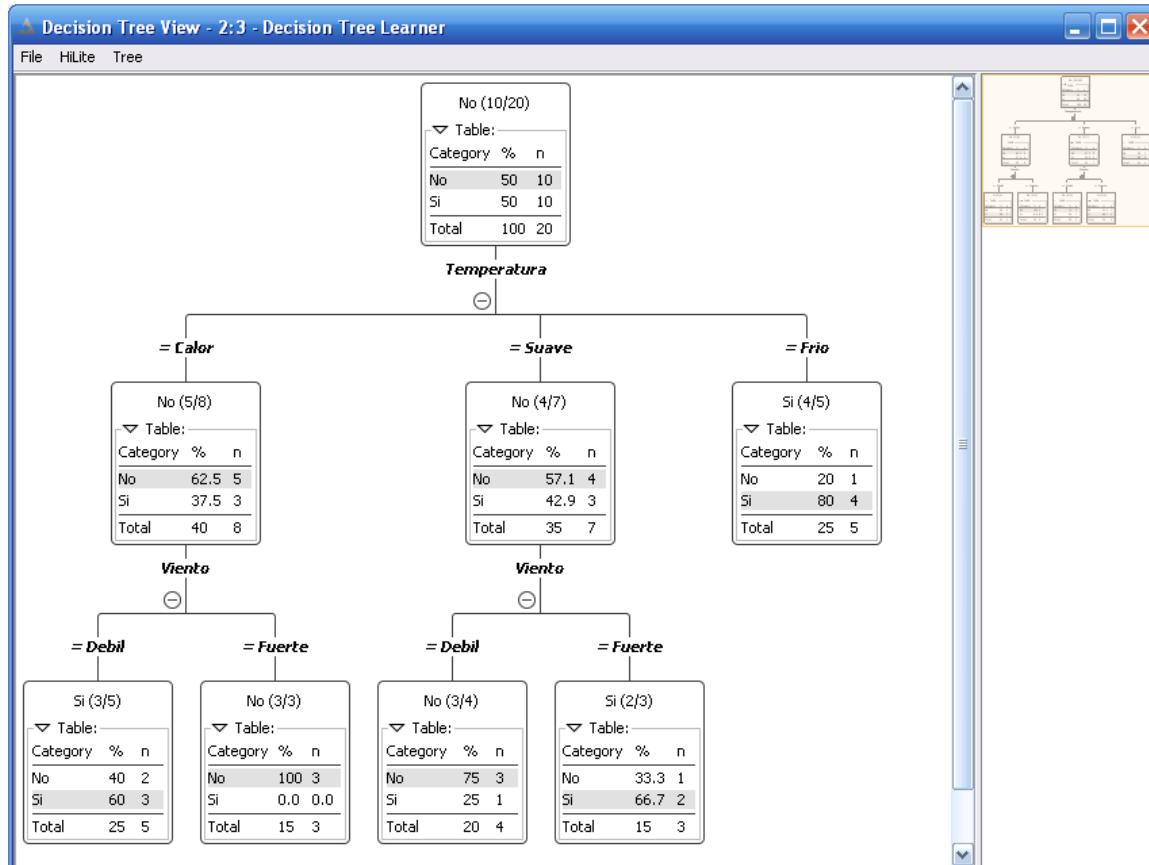
- a) Entrada de datos [IO > Read].
- b) Salida de datos [IO > Write].
- c) Preprocesamiento [Data Manipulation], para filtrar, discretizar, normalizar, filtrar, seleccionar variables...
- d) Minería de datos [Mining], para construir modelos (reglas de asociación, clustering, clasificación, MDS, PCA...).
- e) Salida de resultados [Data Views] para mostrar resultados en pantalla (ya sea de forma textual o gráfica).

Para crear un flujo de ejecución, las salidas de unos nodos se utilizan como entradas de otros. Por ejemplo, un flujo básico podría ser de la forma:



Nodo de lectura de datos

- Nodo de preprocessamiento
- Nodo de modelado (por ejemplo, modelo de clasificación)
- Nodo de salida de resultados.



Lea ahora la introducción a KNIME que se encuentra en la siguiente URL:

<http://tech.knime.org/getting-started>

De forma complementaria, también puede consultar la siguiente presentación de KNIME en castellano:

<http://www.exa.unicen.edu.ar/catedras/dmining/clases/PresentacionKNime.pdf>

IV

ACTIVIDADES

Instalación de KNIME

Descargue la versión de KNIME adecuada para su sistema operativo (Windows o Linux, 32 ó 64 bits): <http://knime.org/download>

You are here: / Home / Downloads / KNIME Desktop

Downloads

- / KNIME Desktop
- / Download
- / Update Site
- / Extensions
- / License
- / Previous Versions
- / Data Sets
- / KNIME Reporting
- / Example Workflows

KNIME Desktop

Proceed directly to download (without registering for release announcements)

Register for release announcements (fill out the form below to get release announcements)

The personal data you enter here will be stored and used for no other reason than to send you messages regarding KNIME updates, bug fixes, and KNIME workshop announcements. Your data will only be disclosed to the entities directly involved with the development and release of KNIME software.

You are here: / Home / Downloads / KNIME Desktop / Overview

Downloads

- / KNIME Desktop
- / Download
- / Update Site
- / Extensions
- / License
- / Previous Versions
- / Data Sets
- / KNIME Reporting
- / Example Workflows

Download KNIME

KNIME Desktop

The KNIME Desktop version 2.3.1 is intended for end users and provides everything needed to immediately begin using KNIME as well as extend KNIME with extension packages developed by others.

Windows

Usually unzipping the archive somewhere on your hard drive is sufficient for the installation of KNIME. However, under Windows problems with the built-in unzip utility sometimes truncate file names. Therefore we offer self extracting archives:

- KNIME for Windows 32bit (self-extracting archive)
- KNIME for Windows 64bit (self-extracting archive)

If you are using a proper unzipper and want to use zip archives instead, you can find them [here](#).

Linux

For Linux a 32 and 64bit build are available:

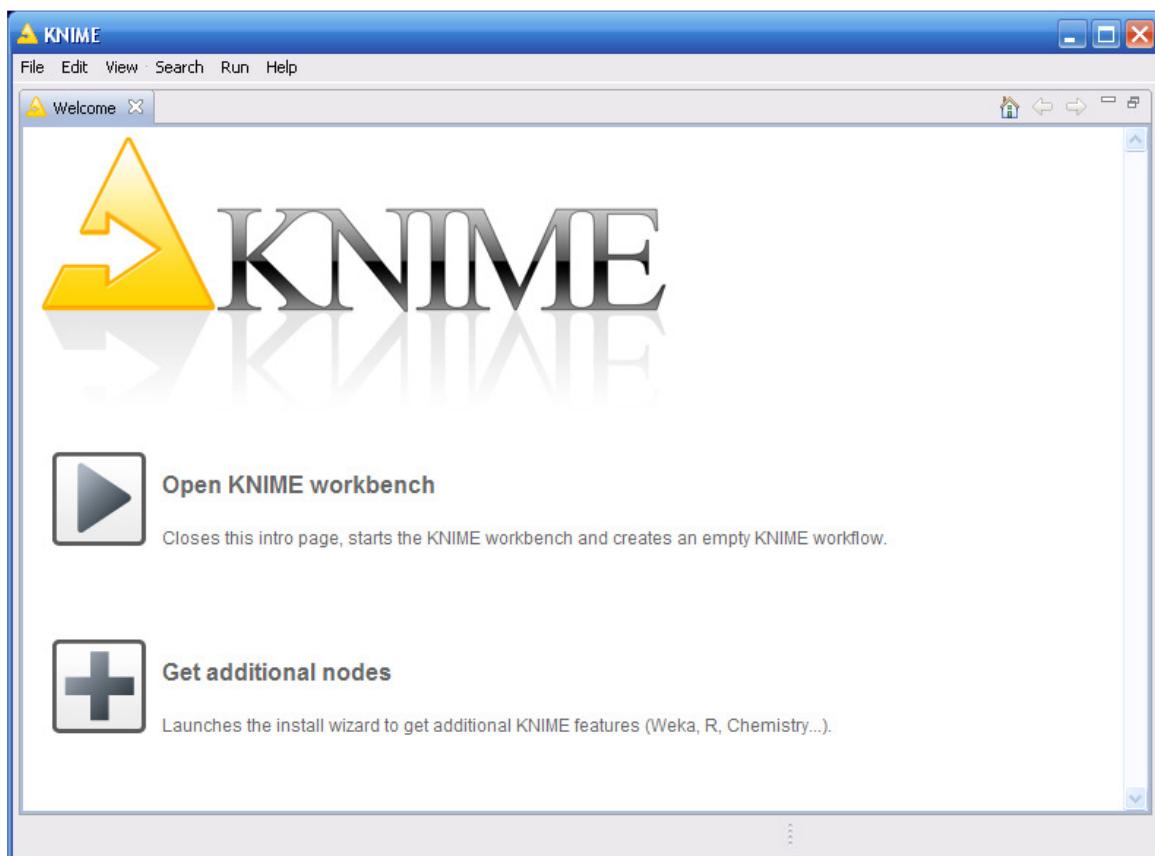
- KNIME for Linux 32bit
- KNIME for Linux 64bit

Si utiliza Windows, ejecute directamente el archivo ZIP autodescomprimible que nos hayamos descargado del sitio oficial de KNIME (p.ej. knime_2.3.1.win32.win32.x86.exe) o descomprima manualmente el archivo en la carpeta en la que deseé instalar el programa.

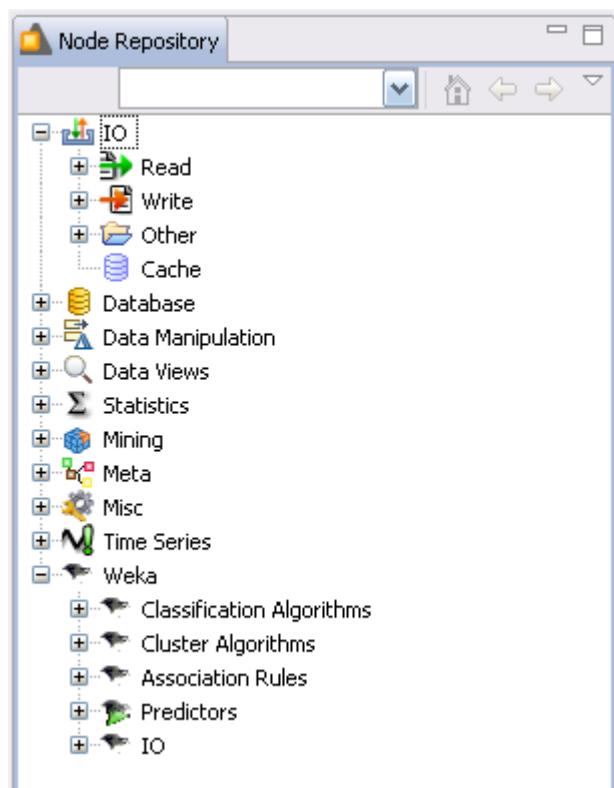
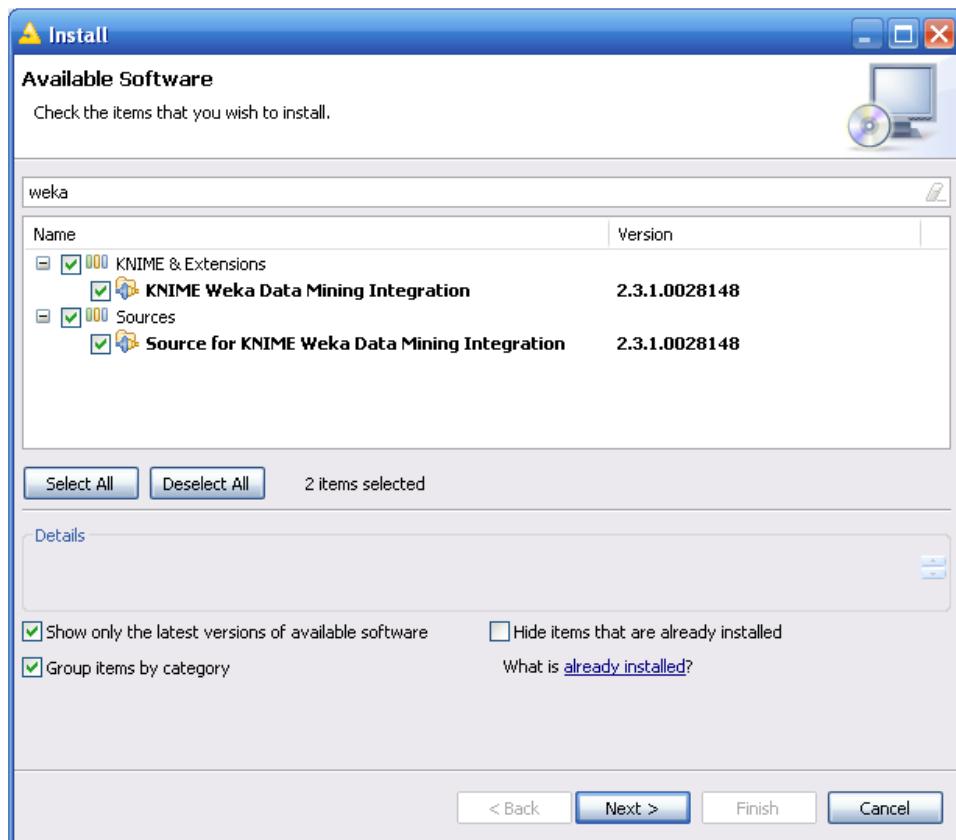
NOTA:

KNIME no requiere programa de instalación. El paquete se desinstala, simplemente, borrando la carpeta en la que lo hayamos descomprimido.

Para ejecutar KNIME, busque el fichero knime.exe en la carpeta en la que haya descomprimido el paquete de instalación y ejecútelo:



Antes de empezar a utilizar KNIME, nos aseguraremos de instalar los componentes de Weka, utilizando la opción “Get additional nodes” de la ventana de inicio de KNIME o accediendo a ellos a través del menú File > Install KNIME Extensions.



Información adicional en la guía de instalación paso a paso:
<http://tech.knime.org/installation-0>

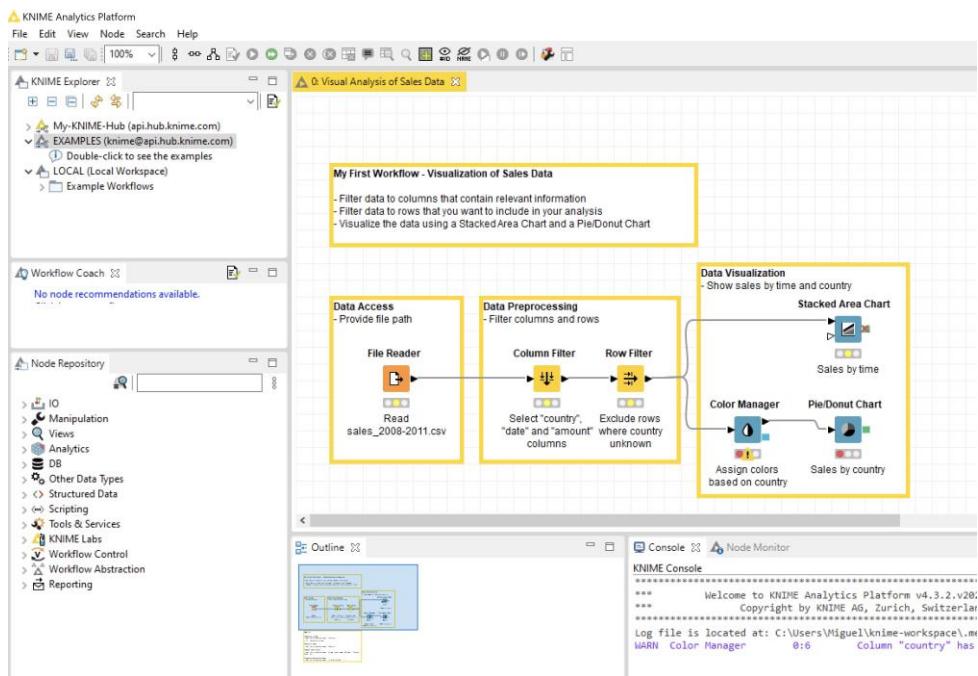
Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

V

EJERCICIOS RESUELTOS

Una vez instalado debe de lucir de la siguiente forma.



VI

EJERCICIOS PROPUESTOS

Proponer una base de datos y desarrollar los mismos pasos que el ejercicio planteado.

VII

CUESTIONARIO

1. Investigar sobre otras herramientas que se pueden utilizar para analizar datos.
2. Hacer una breve descripción de cada herramienta para análisis que hay encontrado.

VIII

BIBLIOGRAFÍA

[1] R. Silipo. KNIME Advanced Luck: A Guide to KNIME Analytics Platform for Advanced Users. KNIME Press, 2018.

Mag. Christian Alain Revilla Arroyo
Mag. Richard Smith Escobedo Quispe

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos



Diseño del Data Warehouse y ETL

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo principal reconocer las herramientas de extracción, modelamiento y carga de datos

II

TEMAS A TRATAR

- Tema 1. Diseño del Data Warehouse y ETL

III

MARCO TEÓRICO

- Pentaho es una plataforma de Business Intelligence (BI) orientada a la solución y centrada en procesos que incluye los componentes requeridos para implementar soluciones basadas en procesos como minería de datos, ETL, generación de informes...
- Actualmente, muchas empresas utilizan Pentaho como herramienta BI para la mejora de la capacidad de análisis y de toma de decisiones, puesto que ofrece soluciones que se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio.

IV

ACTIVIDADES

En este documento se busca realizar el análisis de una base de datos SAKILA que se utilizará como prototipo para el desarrollo de este caso práctico, al cual se le construirá o desarrollará una solución de inteligencia de negocio.

De esta manera se deberá presentar una posible solución de acuerdo a la base de datos entregada, con el fin de poner en práctica todos los conceptos y herramientas estudiadas de inteligencia de negocio para el desarrollo del mismo.

En este proyecto se implementarán herramientas de pentaho para poder llevar a cabo el desarrollo del mismo. Todas las herramientas las podemos encontrar en la siguiente página web: <http://sourceforge.net>

Estas herramientas son:

- ✓ Consola de Pentaho = Creación Cuadros de mando "DashBoard" y Creación de informes
- ✓ kettle o Spoon = Para la creación de los ETL
- ✓ Mondrian de pentaho = Para la realización de los Cubos Olap

MODELO DE NEGOCIO

La empresa la cual se le desarrollará una solución de inteligencia de negocio será CLUBFIML S.A.

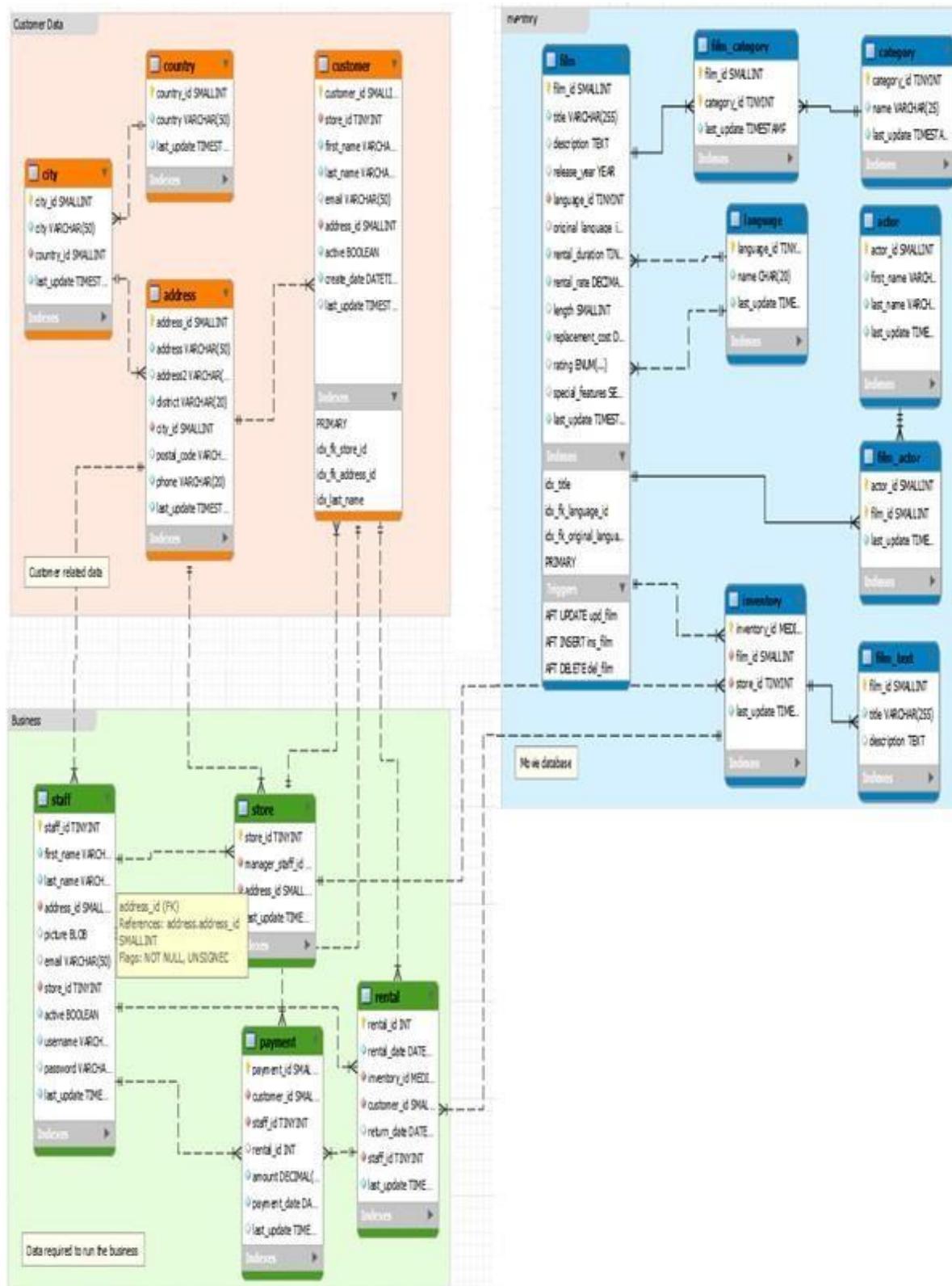
La empresa Sakila tiene como objetivo principal el alquiler de películas de videos, manejando una gran variedad de películas con el fin de satisfacer todas las necesidades de sus clientes.

La empresa Sakila tiene varias sucursales, donde sus clientes podrán encontrar las mejores películas del mercado distribuidas por, lenguaje, actor, título, categoría, de la mejor manera posible, para su fácil acceso dentro de sus tiendas.

El modelo de negocio de la empresa CLUBFILM S.A se compone de los siguientes procesos según el análisis de la base de datos:

- ✓ Datos Clientes: Donde se almacenan los datos de los clientes que por lo menos una vez, han prestado los servicios ofrecidos en las tiendas de CLUBFILM S.A.
- ✓ Inventario: Donde se almacena todos los datos sobre los productos (películas) que se tienen en las sucursales, para la administración y el buen funcionamiento de las tiendas.
- ✓ Proceso del Negocio: En este proceso del negocio es donde se encuentran los datos del negocio.

DIAGRAMA ENTIDAD RELACION "SAKILA"





TABLAS DE HECHOS "MODELO EN ESTRELLA"

1. TABLA "HECHOS SALES"

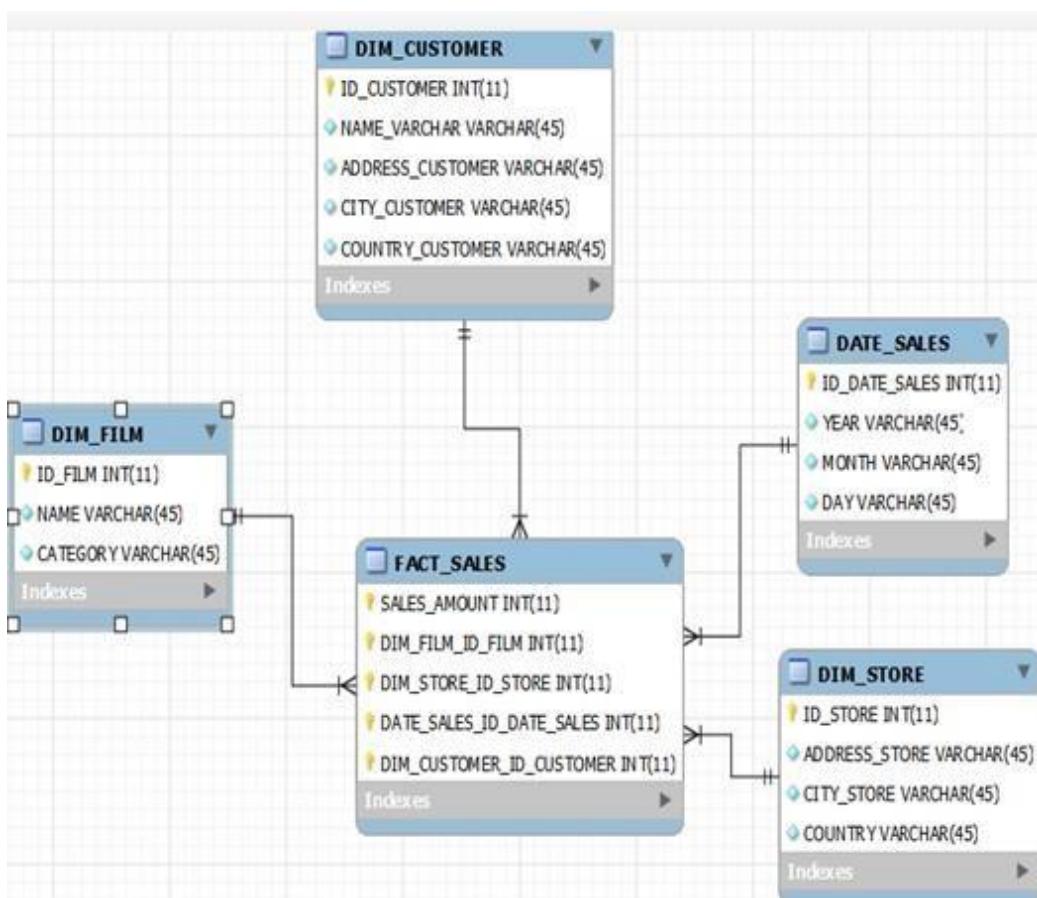


TABLA DE HECHOS	TABLA DE DIMENSIONES
FACT_SALES	DIM_FILM DIM_CUSTOMER DATA_SALES DIM_STORE

2. TABLA "HECHOS INVENTORY"

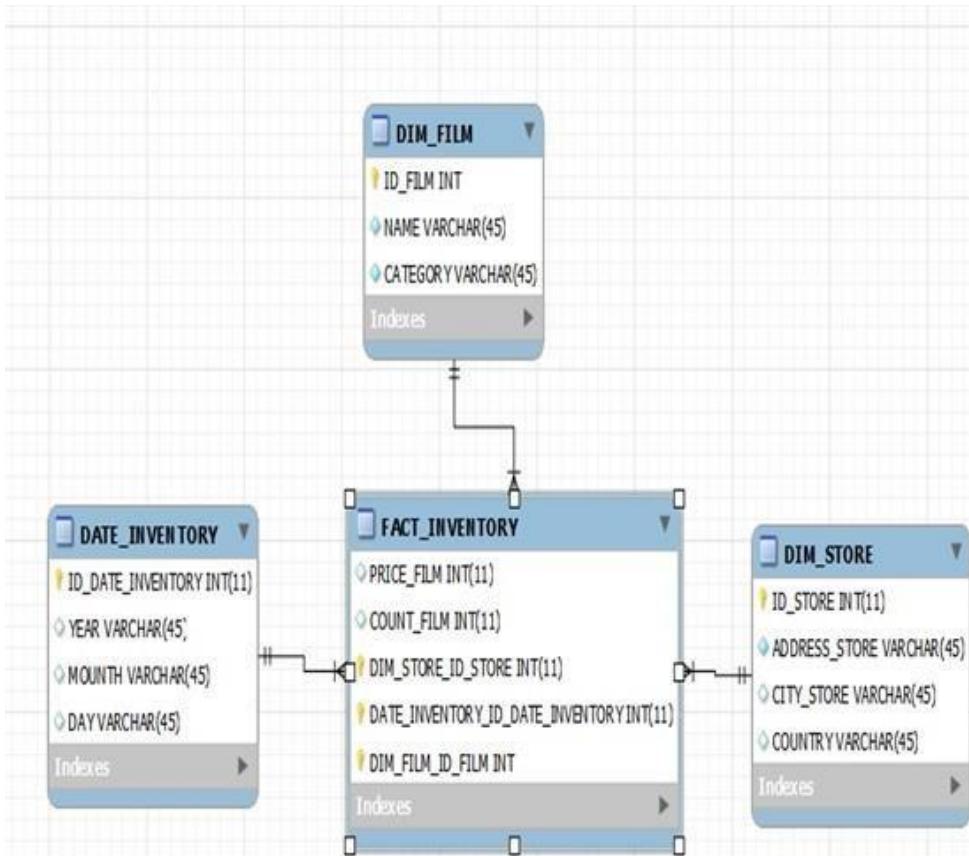


TABLA DE HECHOS	TABLA DE DIMENSIONES
FACT_INVENTORY	DIM_FILM DIM_STORE DATE_INVENTORY

3. TABLA "HECHOS RENTALS"

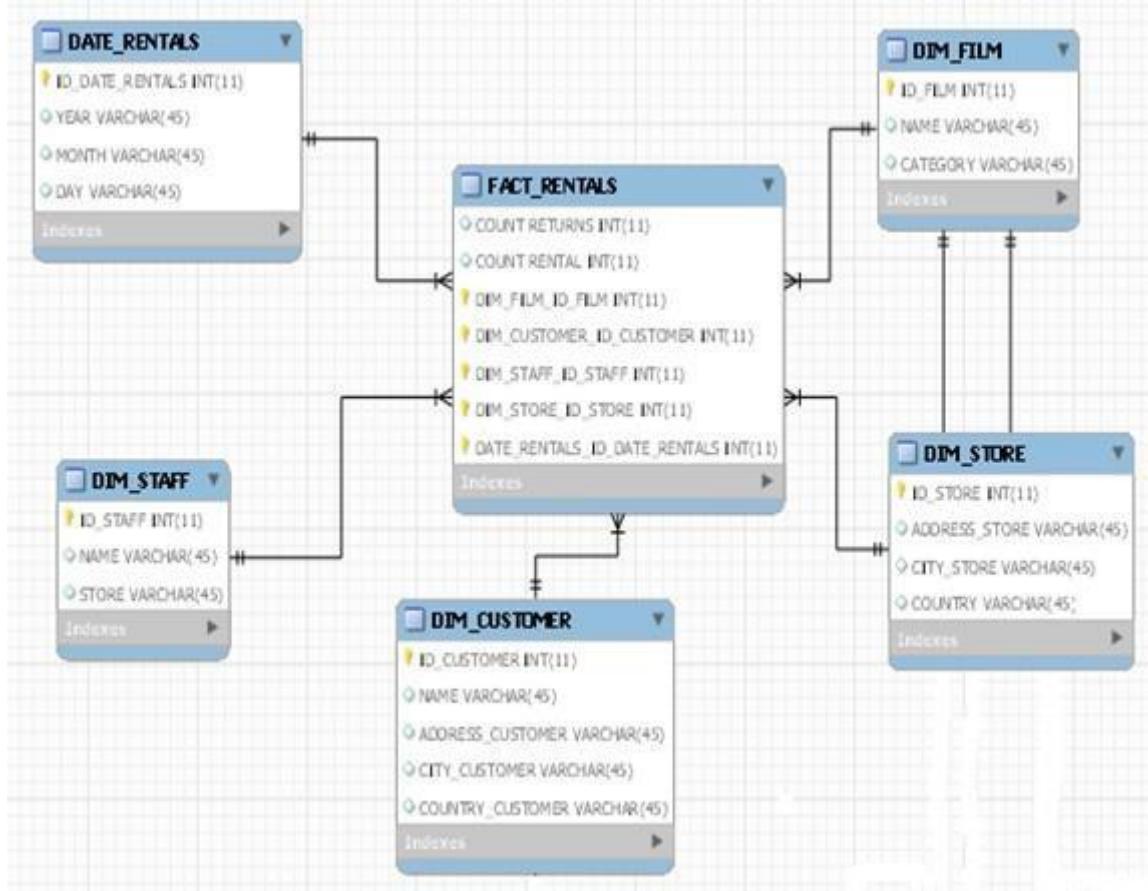


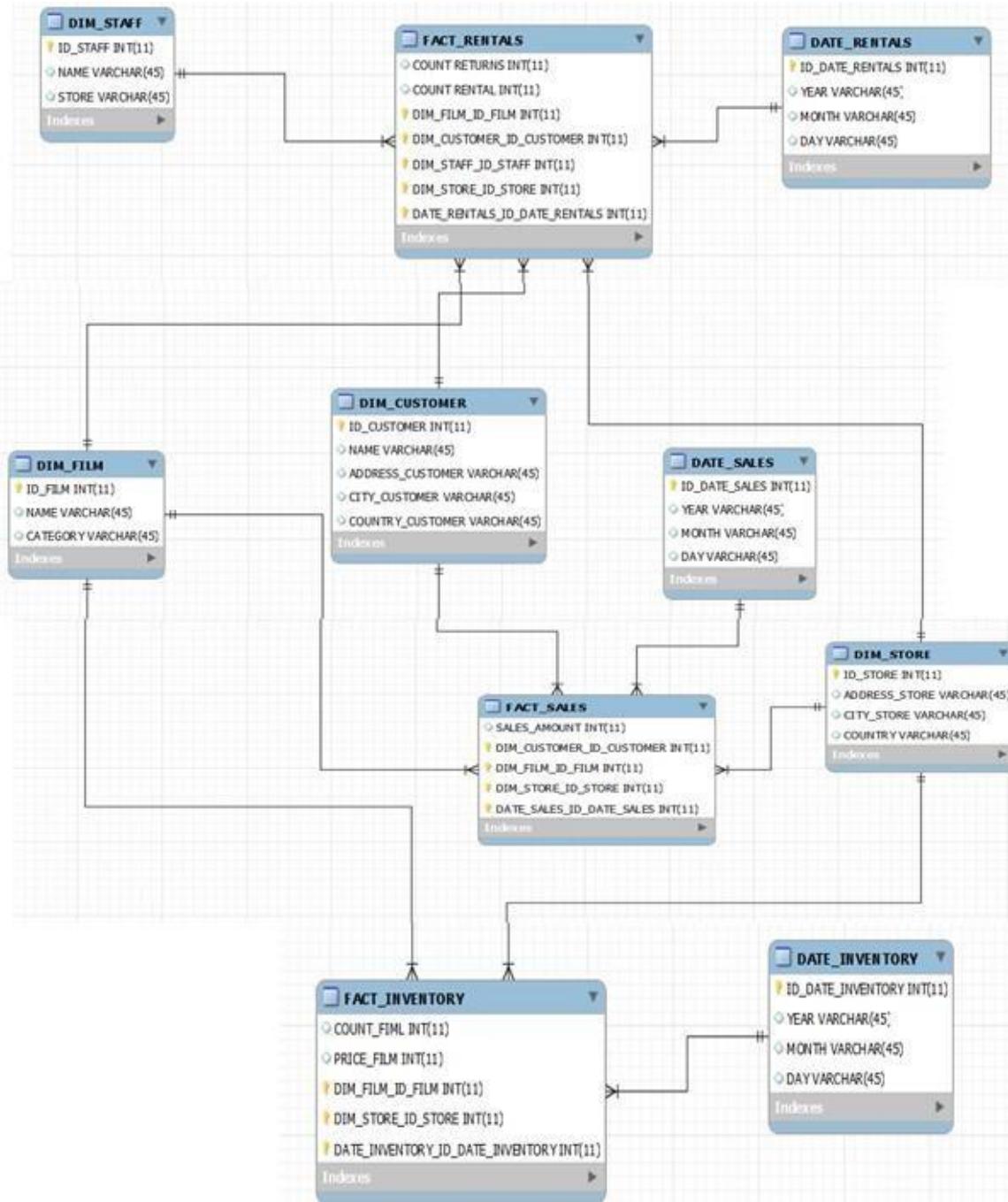
TABLA DE HECHOS

FACT_RENTALS

TABLA DE DIMENSIONES

DIM_FILM
DIM_STORE
DIM_CUSTOMER
DIM_STAFF
DATE_RENTALS

MODELO DIMENSIONAL "DATAWAREHOUSE"



Hasta aquí tenemos nuestro modelo dimensional pero se encuentra vacío, este modelo tiene que diseñarse en la base de datos que mas nos guste manejar, en mi caso la diseño en mysql al igual que la base de datos SAKILA. Ahora debemos crear los ETL's para extraer los datos de la base de datos SAKILA y almacenarla a nuestro nuevo MODELO DIMENSIONAL.

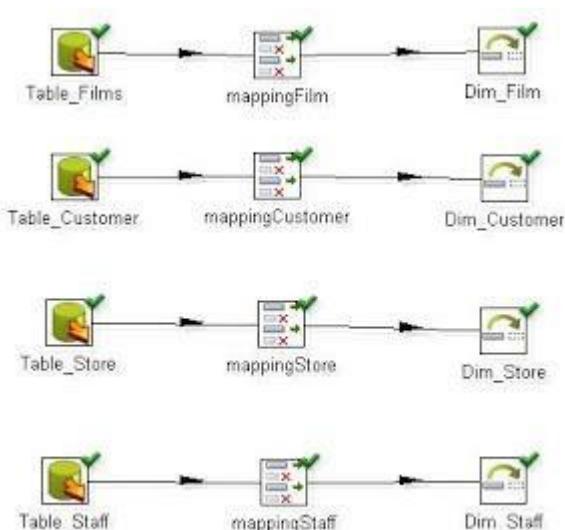
DISEÑO Y CONSTRUCCION DE ETL EN SPOON

Para llevar a cabo el desarrollo de los ETL se contará con la herramienta de PENTAHO llamada SPOON dada para el desarrollo de este proyecto, se debe observar nuestro modelo dimensión presentado anteriormente. Se debe desarrollar un ETL por cada dimensión, fact, y tabla de tiempo que tengamos. La construcción de los ETL nos permitirá extraer los datos de nuestra base de datos (Sakila) y cargarlos en nuestro modelo dimensional desarrollando en el punto anterior.

En nuestra solución decidimos realizar una etl para cada tabla tiempo que posee las tablas de hechos.

ETL DIMENSIONES

A continuación, explicaremos paso a paso la construcción de todos los ETL DIMENSIONES: Dim_films, Dim_customer, Dim_store, Dim_staff



Se realizarán los mismos pasos para cada uno de los ETL

Paso 1: Primero construimos una tabla de entrada en SPOON la cual le asignamos el nombre y determinamos la conexión a la base de datos y en sql determinamos que tabla y que columnas deseamos que nos extraiga y nos muestre, damos PREVISUALIZAR para observar los datos.

Paso 2: Construimos una selecciona/renombrar valores para determinar que valores me renombrar de la base de datos y el modelo dimensional para luego le los almacene en los respectivos campos.

Paso 3: por ultimo construimos un insertar/actualizar para insertarle los datos extraídos de la base de datos al modelo dimensional.

Construcción de Dimensión de Tiempo:

Es redundante tener 3 dimensiones de tiempo en nuestro DataWarehouse, no creería que sería mejor tener una sola dimensión de tiempo que contenga todas las posibles fechas de nuestras Fact Tables ? Esta seria una mejor que pueden implementar en sus soluciones.

Con el Query a continuación podemos obtener todas esas fechas de nuestras Fact Tables, solo quería construir el campo incrementar para hacer único cada uno de los registros que contiene nuestra Dim_Tiempo y por el cual será la relación con las Fact:

Dim_Tiempo:

```

Select date(a.rental_date) AS Fecha,
       year(a.rental_date) AS Año,
       quarter(a.rental_date) AS Trimestre,
       Month(a.rental_date) AS Mes,
       monthname(a.rental_date) AS
       MesNombre, Day(a.rental_date) AS Dia
  From
  
```

sakila.rental a

Union All

Select

```

      date(b.payment_date) AS Fecha,
      year(b.payment_date) AS Año,
      quarter(b.payment_date) AS Trimestre,
      Month(b.payment_date) AS Mes,
      monthname(b.payment_date) AS
      MesNombre, Day(b.payment_date) AS Dia
  From
  
```

sakila.payment b

Union All

Select

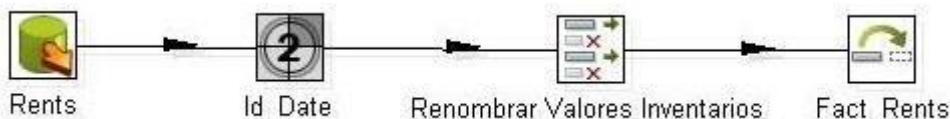
```

      Date(c.last_update) AS Fecha,
      year(c.last_update) AS Año,
      quarter(c.last_update) AS Trimestre,
      Month(c.last_update) AS Mes,
      monthname(c.last_update) AS MesNombre,
      Day(c.last_update) AS Dia
  From sakila.inventory c;
  
```

ETL HECHOS

Para la construcción de los ETL de las tablas de hechos, realizaremos cuatro pasos, los mismos pasos son para cada uno de las tablas.

Tomaremos como ejemplo la tabla hechos rents



Paso 1: Primero construimos una tabla de entrada en SPOON la cual le asignamos el nombre y determinamos la conexión a la base de datos y en sql determinamos que tabla y que columnas deseamos que nos extraiga y nos muestre, damos PREVISUALIZAR para observar los datos.

Paso 2: Construimos una selecciona/renombra valores para obtener el valor de la secuencia de la base de datos

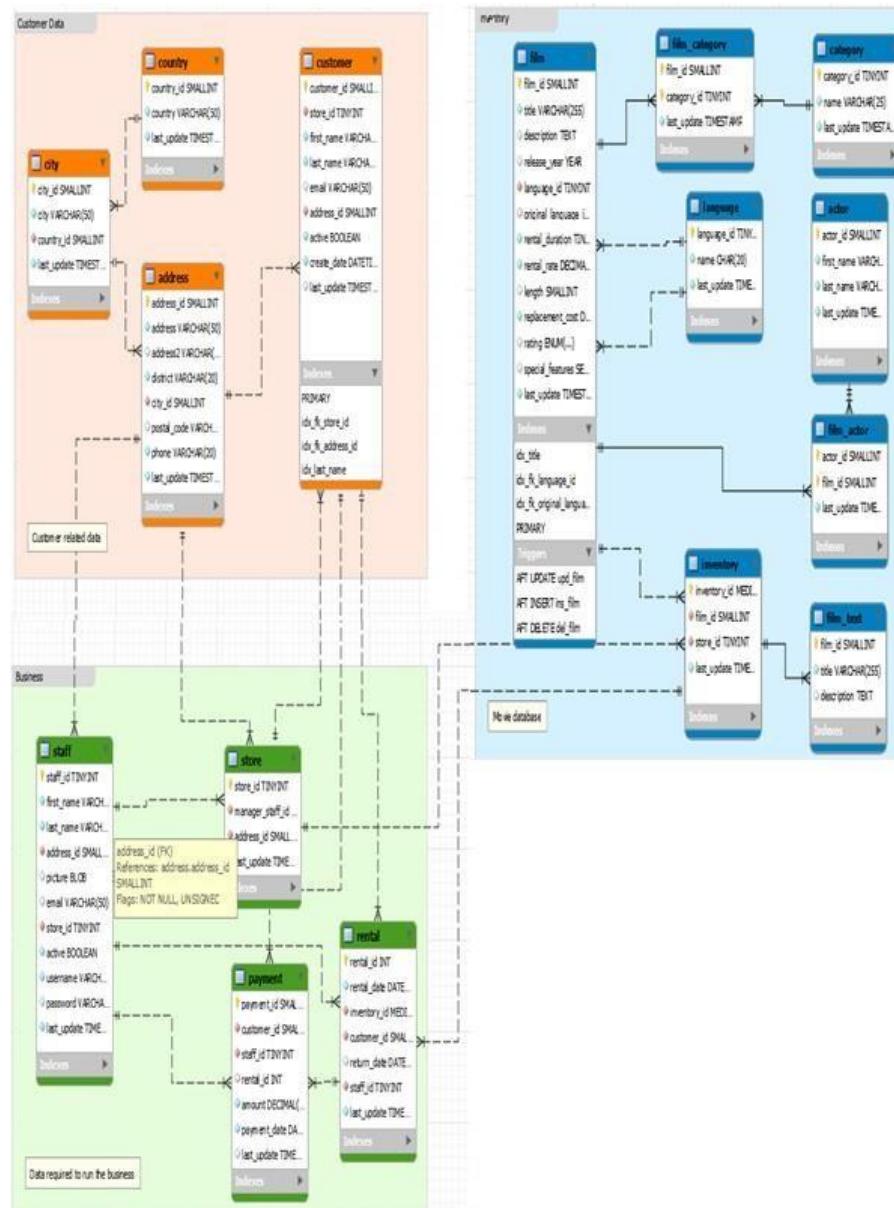
Paso 3: Construimos una selección/renombra valores para determinarle que valores me renombra de la base de datos y el modelo dimensional para luego le los almacene en los respectivos campos

Paso 4: por último, construimos un insertar/actualizar para insertarle los datos extraídos de la base de datos al modelo dimensional.

V

EJERCICIOS RESUELTOS

Una vez desarrollado el ejercicio deberá de lucir de la siguiente forma.



Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

VI

EJERCICIOS PROPUESTOS

- En base al ejemplo visto deberá de desarrollar un ejercicio, usted tendrá la libertad de proponer las tablas y las relaciones.
- Deberá también de describir los componentes de su diagrama ETL

VII

CUESTIONARIO

1. ¿Cuál es la utilidad de usar ETL?
2. ¿Qué otras herramientas puedo utilizar para generar ETL's?.

VIII

BIBLIOGRAFÍA

[1] A. Meadows y M. Roldán. Pentaho Data Integration. 2nd. ed.. Cookbook. Packt Publishing, 2013.

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos



Minería de Datos

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo principal el que el estudiante se familiarice con la herramienta R.

II

TEMAS A TRATAR

- Tema 1. Minería de Datos

III

MARCO TEÓRICO

- R es el entorno de programación y un lenguaje de popular utilización en el ámbito de la estadística. Una de las principales ventajas de este programa es su gratuidad. Es similar a otros paquetes de programación como Matlab (no gratuito), y mucho más simple que otros lenguajes como C++ o Fortran. Se puede utilizar R en su plataforma original, pero por motivos educacionales es preferible trabajar en R desde R-studio (también gratuito), por disponer este último de ventanas mejor organizadas y contar con más opciones gráficas.

IV

ACTIVIDADES

Actividad 1.

- ✓ Seguir los pasos siguientes para la instalación de R y otros elementos necesarios para el análisis de datos.

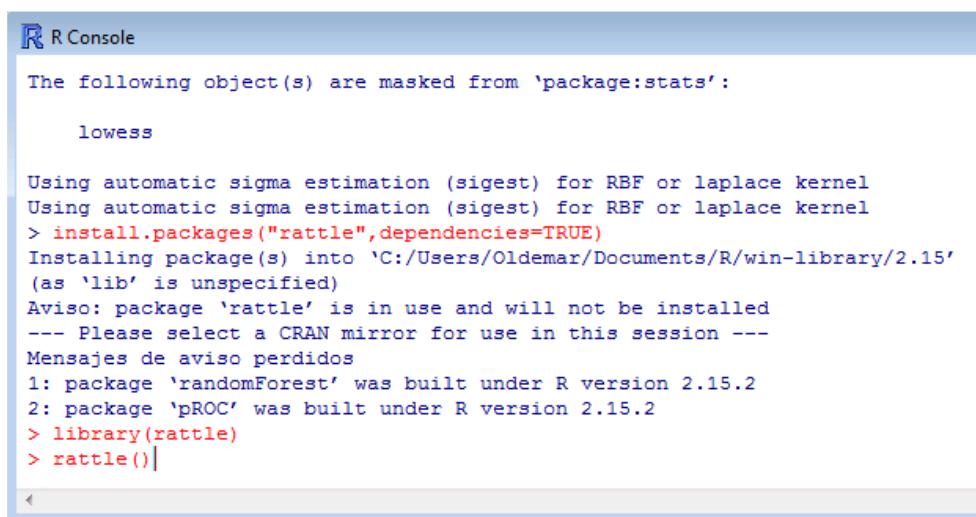
Instalación de R, RStudio, Rattle y de FactorMineR

Instalando R

- Puede bajarse en <http://www.r-project.org>
- Directamente se puede bajar de: <http://lib.stat.cmu.edu/R/CRAN/> e instálelo ejecutando:
 - R-2.15.1-win.exe
- Instalar RStudio ejecutando:
 - RStudio-0.96.331.exe

Rattle

- Para instalarlo:
 - `install.packages("rattle",dependencies=TRUE)`
- Para ejecutarlo:
 - `library(rattle)`
 - `rattle()`
- Sitio WEB:
 - <http://rattle.togaware.com/>



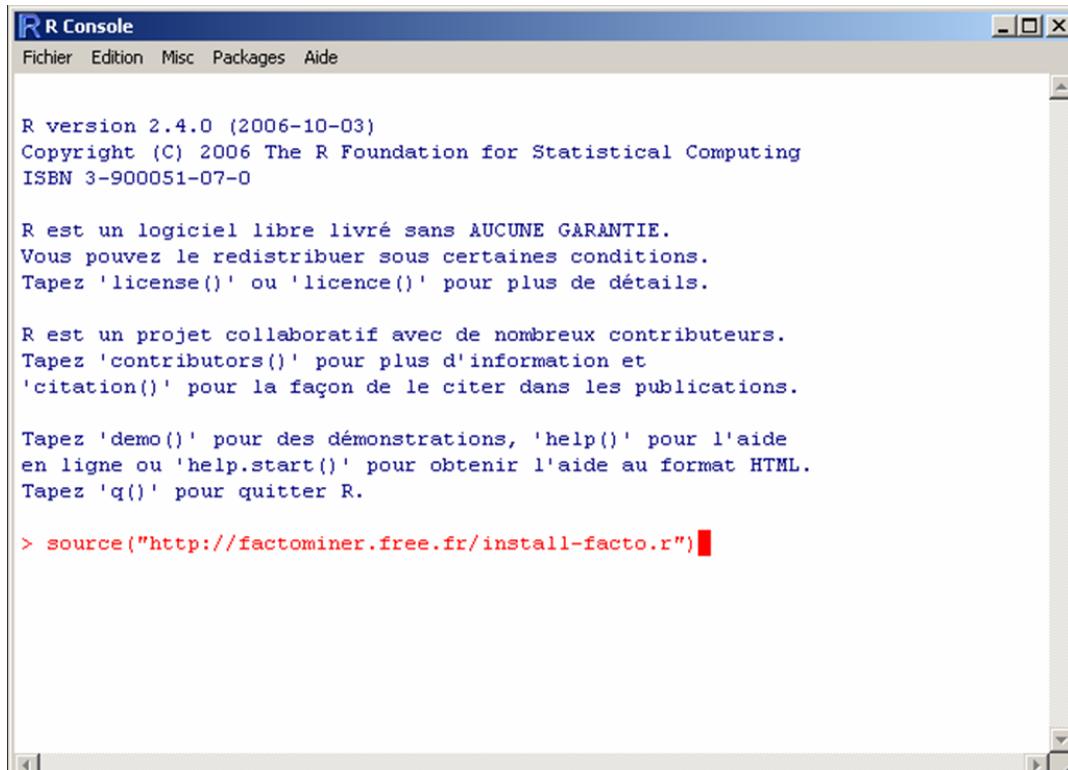
```
R Console
The following object(s) are masked from 'package:stats':
  lowess

Using automatic sigma estimation (sigest) for RBF or laplace kernel
Using automatic sigma estimation (sigest) for RBF or laplace kernel
> install.packages("rattle",dependencies=TRUE)
Installing package(s) into 'C:/Users/Oldemar/Documents/R/win-library/2.15'
(as 'lib' is unspecified)
Aviso: package 'rattle' is in use and will not be installed
--- Please select a CRAN mirror for use in this session ---
Mensajes de aviso perdidos
1: package 'randomForest' was built under R version 2.15.2
2: package 'pROC' was built under R version 2.15.2
> library(rattle)
> rattle()|
```

Instalando el “FactoMineR GUI”

- Ejecute el software R.
- Estando conectado a Internet copie, pegue y ejecute la siguiente línea en la barra de comandos de R, como se ve en la pantalla de la siguiente filmina:
`source("http://factominer.free.fr/install-facto.r")`

Instalando “FactoMineR GUI”



R version 2.4.0 (2006-10-03)
Copyright (C) 2006 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

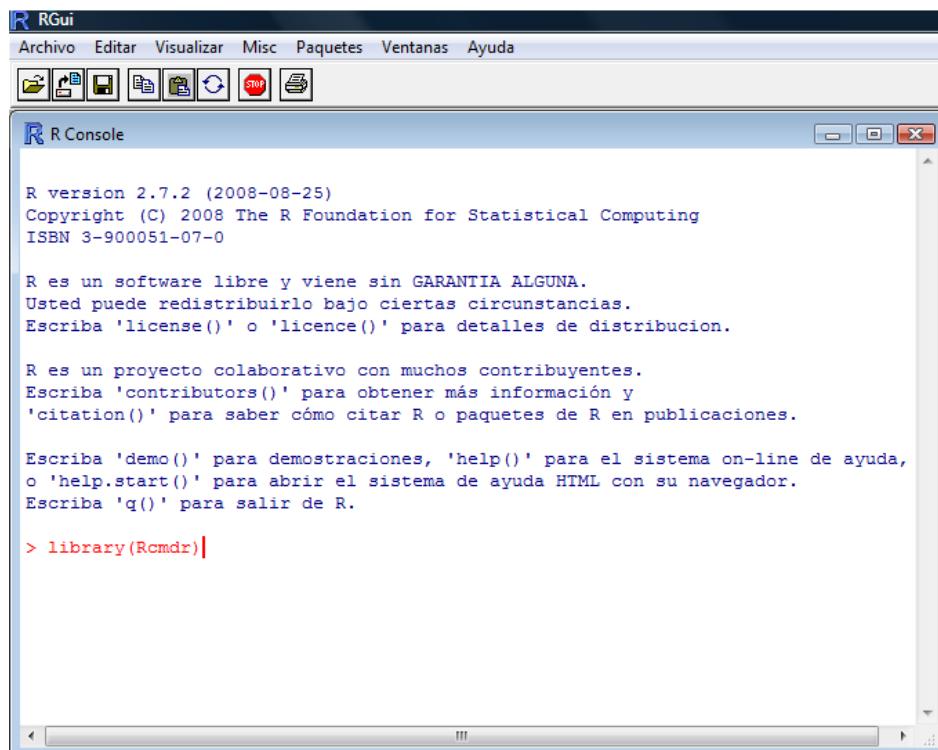
```
> source("http://factominer.free.fr/install-facto.r")
```

Usando “FactoMineR GUI”

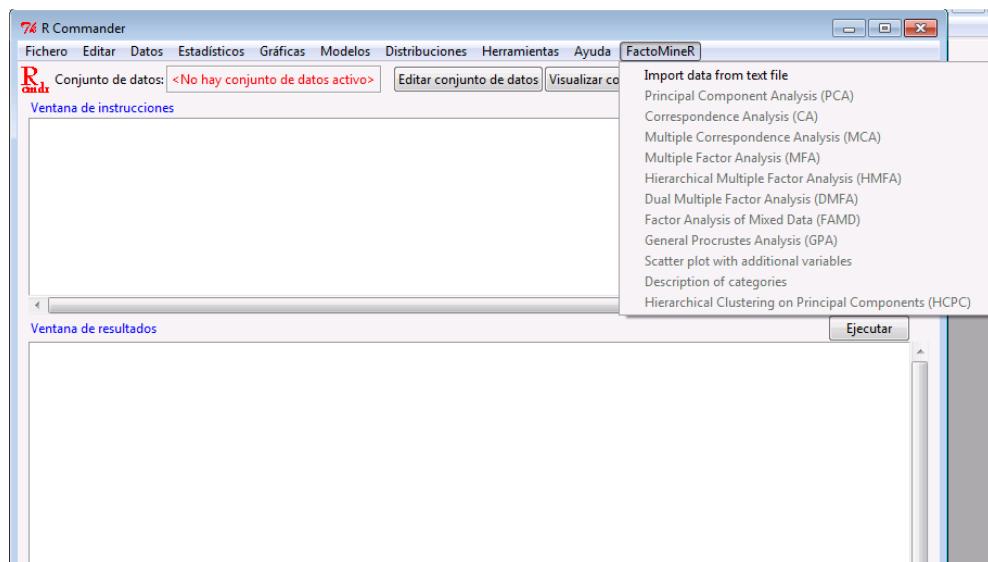
Para verificar que FactoMiner quedó bien instalado ejecute desde R el siguiente comando, como se muestra en la siguiente filmina:

- library(Rcmdr)

Usando “FactoMineR GUI”



Deberá aparecerle el siguiente Menú:



- ✓ Para el caso de la instalación de la herramienta FactorMineR (en el caso que no funcione en el paso anterior), deberá de usar como guía el siguiente link:
<https://www.youtube.com/watch?v=mKwgi7CVPvs>

Actividad 2.

- ✓ Realizar la implementación del siguiente ejemplo de Regresión
<https://www.youtube.com/watch?v=ZURs9QfnQuQ>
- ✓ Finalmente deberá de implementar u ejercicio que muestre la interrelación de dos variables a través de u gráfico usado Rattle y/o FactorMineR

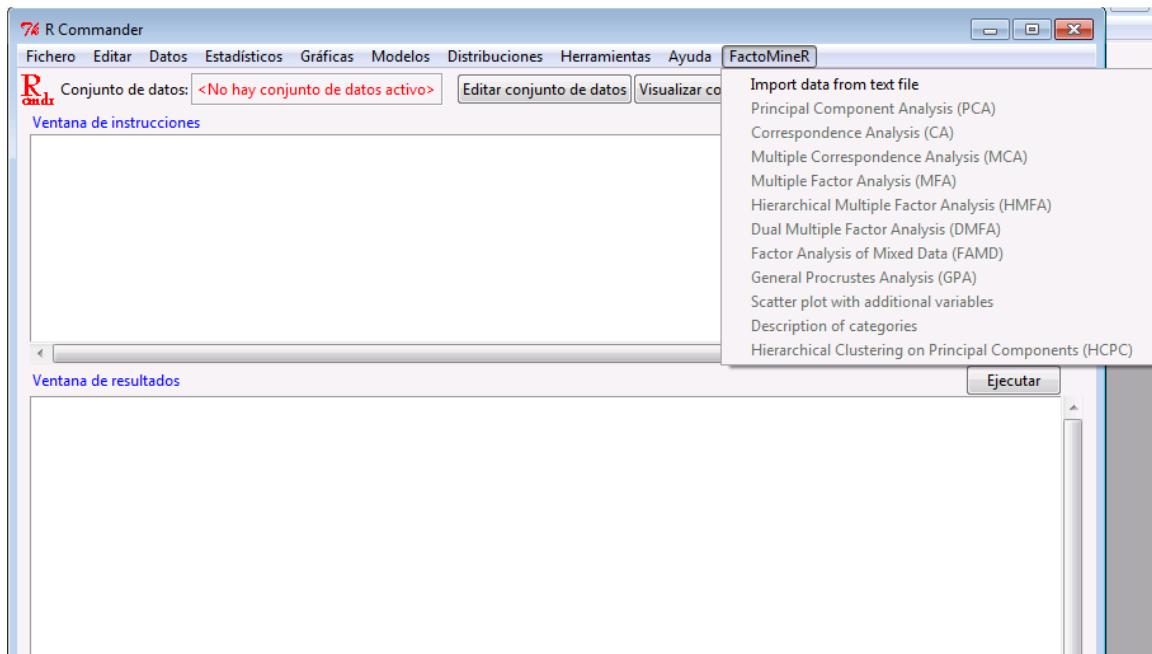
Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

V

EJERCICIOS RESUELTOS

Una vez instalado debe de lucir de la siguiente forma.



VI

EJERCICIOS PROPUESTOS

Deberá de desarrollar en clase un ejercicio que permita realizar el análisis y que le permita evaluar un grupo de variables similar al ejemplo propuesto.

VII

CUESTIONARIO

1. Revisar la herramienta FactoMineR, seleccionar un grupo de datos y hacer uso de la opción de análisis por correspondencia.

VIII

BIBLIOGRAFÍA

[1] H. Wickham y G. Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (English Edition). O'Reilly Media, 2017.

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos

Laboratorio

4

R y métodos de análisis

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo el que el alumno lleve a cabo el análisis de series temporales básico usando el software estadístico RStudio.

II

TEMAS A TRATAR

- Tema 1. R y métodos de análisis

III

MARCO TEÓRICO

- R es el entorno de programación y un lenguaje de popular utilización en el ámbito de la estadística. Una de las principales ventajas de este programa es su gratuidad. Es similar a otros paquetes de programación como Matlab (no gratuito), y mucho más simple que otros lenguajes como C++ o Fortran. Se puede utilizar R en su plataforma original, pero por motivos educacionales es preferible trabajar en R desde R-studio (también gratuito), por disponer este último de ventanas mejor organizadas y contar con más opciones gráficas.

IV

ACTIVIDADES

Actividad 1.

- Ahora que cuenta con la herramienta R instalada, deberá de hacer el análisis de series temporales. <https://www.youtube.com/watch?v=O0DLO0oWd5c>

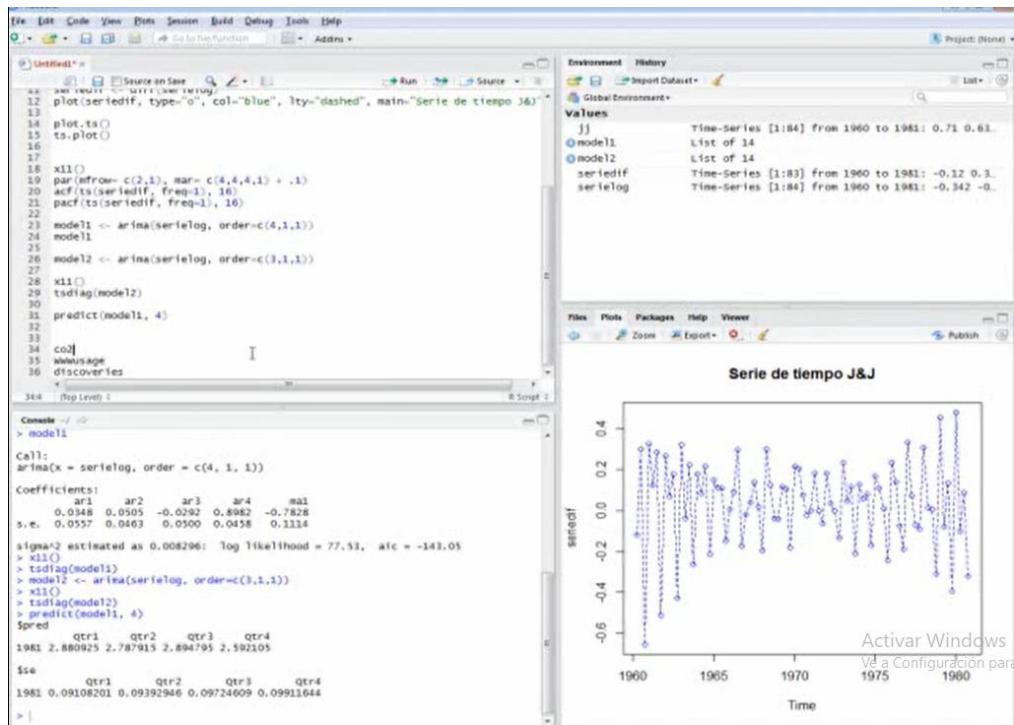
Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

V

EJERCICIOS RESUELTOS

Una vez instalado debe de lucir de la siguiente forma.



VI

EJERCICIOS PROPUESTOS

Proponer una base de datos y desarrollar los mismos pasos que el ejercicio planteado para series temporales.

VII

CUESTIONARIO

1. Indicar los tipos de aplicaciones sobre los que se puede trabajar usando series temporales.

VIII

BIBLIOGRAFÍA

[1] H. Wickham y G. Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (English Edition). O'Reilly Media, 2017.

Mag. Christian Alain Revilla Arroyo
Mag. Richard Smith Escobedo Quispe

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos



Clustering

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo el que el alumno lleve a cabo el análisis del modelo clustering usando herramientas como Python y sus librerías gráficas.

II

TEMAS A TRATAR

- Tema 1. Clustering

III

MARCO TEÓRICO

- El Clustering es una tarea que consiste en agrupar un conjunto de objetos (no etiquetados) en subconjuntos de objetos llamados Clusters. Cada Cluster está formado por una colección de objetos que son similares (o se consideran similares) entre sí, pero que son distintos respecto a los objetos de otros Clusters.
- Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional.

IV

ACTIVIDADES

Actividad 1.

Revisar el siguiente ejemplo en el que se explica un ejemplo de clustering usando Python. Deberá de instalar los elementos necesarios para que funcione la implementación

Clustering con Python

https://www.youtube.com/watch?v=VK_2_dxSTwA

Nota: La base de datos de iris.csv la puede tomar como referencia de cualquiera de los siguientes links.

<https://gist.github.com/netj/8836201>

<https://www.kaggle.com/saurabh00007/iriscsv>

Rúbrica de evaluación

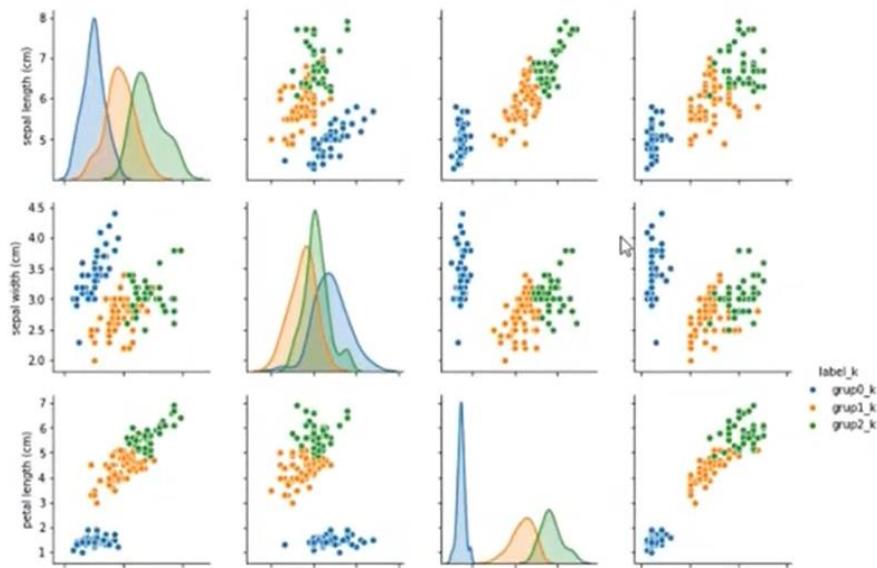
Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

Mag. Christian Alain Revilla Arroyo

Mag. Richard Smith Escobedo Quispe

V**EJERCICIOS RESUELTOS**

Una vez desarrollado el ejercicio debe de lucir de la siguiente forma.



VI**EJERCICIOS PROPUESTOS**

Proponer una base de datos y desarrollar los mismos pasos que el ejercicio planteado.

VII**CUESTIONARIO**

1. Investigar sobre el algoritmo de clusterización y proponer un ejercicio similar

VIII**BIBLIOGRAFÍA**

- [1] K. Jolly. Machine Learning with scikit-learn Quick Start Guide: Classification, regression, and clustering techniques in Python. Packt Publishing Press, 2018.

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos

Laboratorio
6

Métodos avanzados de Minería de Datos - Validación Cruzada

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo el que el alumno conozca sobre los métodos avanzados sobre minería de datos, en este caso la validación cruzada

II

TEMAS A TRATAR

- Tema 1. Métodos avanzados de Minería de Datos - Validación Cruzada

III

MARCO TEÓRICO

- Validación cruzada: puede aplicarse para estimar el test error asociado a un determinado método de aprendizaje estadístico (tanto regresión como clasificación) para evaluar el rendimiento del modelo (model assessment), pero también para seleccionar niveles apropiados de flexibilidad, como el grado de polinomio, etc. (model selection), usando el test error rate.
- Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional.

IV

ACTIVIDADES

Actividad 1.

Revisar el siguiente ejemplo en el que se explica un ejemplo de validación cruzada usando Python. Deberá de instalar los elementos necesarios para que funcione la implementación

<https://www.youtube.com/watch?v=xq05Uvv7thQ>

Mag. Christian Alain Revilla Arroyo

Mag. Richard Smith Escobedo Quispe

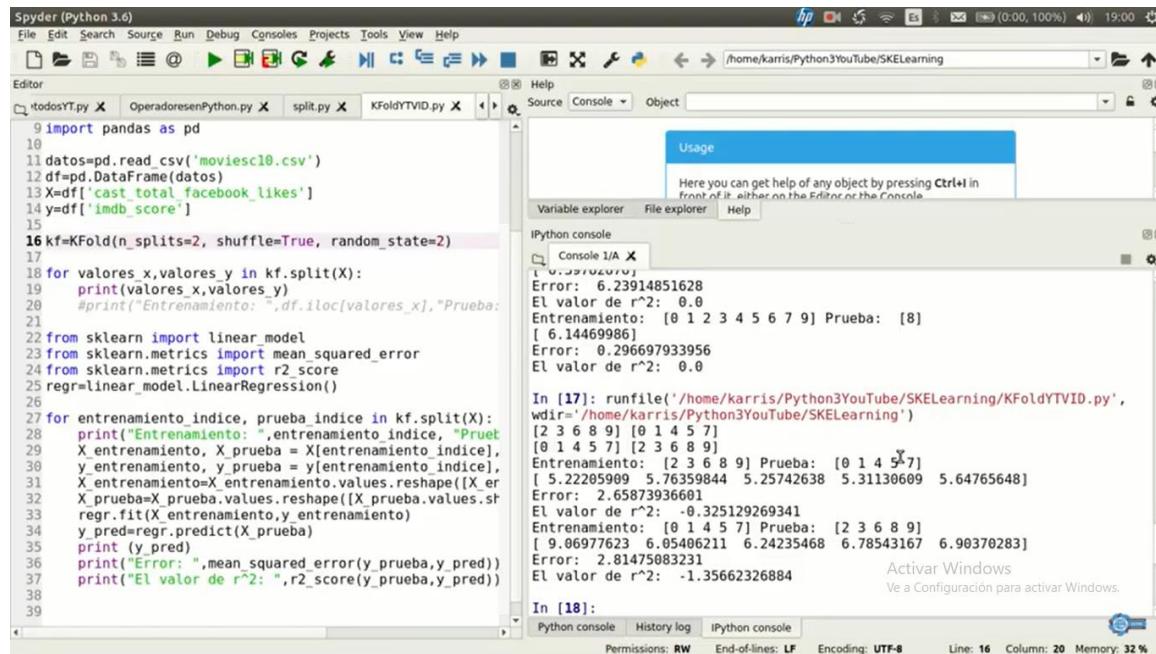
Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			
Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

V

EJERCICIOS RESUELTOS

Finalmente, los resultados podrían lucir de la siguiente forma.



The screenshot shows the Spyder Python 3.6 IDE interface. The code editor contains a script named `KFoldYTVID.py` with Python code for performing K-fold cross-validation on a dataset. The console window displays the execution of the script, showing the training and testing sets, the calculation of the mean squared error, and the resulting r^2 score for each fold. The output is as follows:

```
File Edit Search Source Run Debug Consoles Projects Tools View Help
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor
todosYT.py X OperadoresenPython.py X split.py X KFoldYTVID.py X Help
Source Console Object
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Variable explorer File explorer Help
IPython console
In [17]: runfile('/home/karris/Python3YouTube/SKELearning/KFoldYTVID.py', wdir='/home/karris/Python3YouTube/SKELearning')
[2 3 6 8 9] [0 1 4 5 7]
[0 1 4 5 7] [2 3 6 8 9]
Error: 6.23914851628
El valor de r^2: 0.0
Entrenamiento: [0 1 2 3 4 5 6 7 9] Prueba: [8]
[6.14469986]
Error: 0.296697933956
El valor de r^2: 0.0

In [18]:
Python console History log IPython console
Permissions: RW End-of-lines: LF Encoding: UTF-8 Line: 16 Column: 20 Memory: 32 %
Activar Windows
Ve a Configuración para activar Windows.
In [18]:
```

VI

EJERCICIOS PROPUESTOS

Proponer una base de datos y desarrollar los mismos pasos que el ejercicio planteado.

VII

CUESTIONARIO

1. Indicar cuales serían las ventajas de la validación cruzada sobre otro método.
2. Buscar otros métodos alternativos a la validación cruzada

VIII

BIBLIOGRAFÍA

- [1] K. Jolly. Machine Learning with scikit-learn Quick Start Guide: Classification, regression, and clustering techniques in Python. Packt Publishing Press, 2018.

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



MAG. CHRISTIAN ALAIN REVILLA ARROYO
MAG. RICHART SMITH ESCOBEDO QUISPE

GUÍA DE LABORATORIO

INTELIGENCIA DE NEGOCIOS
2019-B
SEMESTRE X

COMPETENCIAS

Conocimiento de los temas de actualidad.

La capacidad de utilizar las técnicas, habilidades y herramientas modernas de ingeniería y computación necesarias para la práctica de la ingeniería del software.

Habilidad para trabajar en uno o más dominios de aplicación significativos

Laboratorio

7

Deep Learning

I

OBJETIVOS

- Objetivo 1. La presente práctica de laboratorio tiene como objetivo el que el alumno conozca de manera introductoria los conceptos sobre Deep Learning.
-

II

TEMAS A TRATAR

- Tema 1. Deep Learning

III

MARCO TEÓRICO

- Deep Learning. El Deep Learning lleva a cabo el proceso de Machine Learning usando una red neuronal artificial que se compone de un número de niveles jerárquicos. En el nivel inicial de la jerarquía la red aprende algo simple y luego envía esta información al siguiente nivel. El siguiente nivel toma esta información sencilla, la combina, compone una información algo un poco más compleja, y se lo pasa al tercer nivel, y así sucesivamente.
- Continuando con el ejemplo del gato, el nivel inicial de una red de Deep Learning podría utilizar las diferencias entre las zonas claras y oscuras de una imagen para saber dónde están los bordes de la imagen. El nivel inicial pasa esta información al segundo nivel, que combina los bordes construyendo formas simples, como una línea diagonal o un ángulo recto. El tercer nivel combina las formas simples y obtiene objetos más complejos como óvalos o rectángulos. El siguiente nivel podría combinar los óvalos y rectángulos, formando barbas, patas o colas rudimentarias. El proceso continúa hasta que se alcanza el nivel superior en la jerarquía, en el cual la red aprende a identificar gatos.

IV

ACTIVIDADES

Actividad 1.

Revisar el siguiente video en el que se explica el tema “Deep Learning With Python | Deep Learning Tutorial For Beginners” usando Python. Deberá de instalar los elementos necesarios para que funcione la implementación

<https://www.youtube.com/watch?v=X-FKcenZ-jo>

Rúbrica de evaluación

Orden			
Excelente: El trabajo maneja una estructura mejor a la solicitada	Bueno: El trabajo maneja una estructura definida y se ve un correcto orden	Regular: La entrega es deficiente, con conserva una estructura	Deficiente: No presenta la tareas propuesta
4	3	2	0
Aplicación de Temas			
Excelente: Sobrepasa las expectativas, cumple con los enseñado en clase y otros recursos adicionales	Bueno: Aplica todos los conceptos desarrollados en clase	Regular: Aplica los elementos solicitados de manera parcial	Deficiente: No presenta la tareas propuesta
4	3	2	0
Calidad del Contenido			
Excelente: Sobrepasa las expectativas	Bueno: Los contenidos desarrollados no se están resumidos.	Regular: Aplica los elementos solicitados de manera parcial y se o tiene congruencia lo puesto en el trabajo	Deficiente: No presenta la tareas propuesta
4	3	2	0
Creatividad			
Excelente: Sobrepasa las expectativas	Bueno: Cumple con los elementos necesarios con variedad en la presentación y usando correctamente los recursos	Regular: Cumple con los elementos necesarios con variedad en la presentación y uso poco variado los recursos	Deficiente: No presenta la tareas propuesta
4	3	2	0
Cumplimiento			

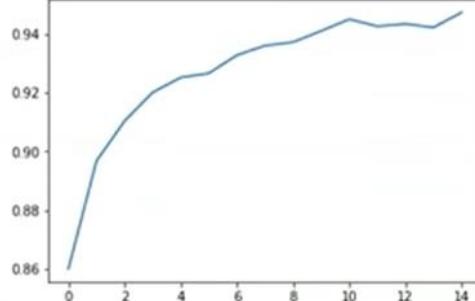
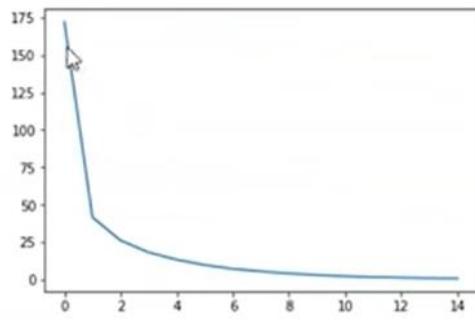
Mag. Christian Alain Revilla Arroyo

Mag. Richard Smith Escobedo Quispe

Excelente: Presenta el mismo día su práctica	Bueno: Entrega dentro de plazo establecido	Regular: Presenta fuera del plazo establecido	Deficiente: No presenta la tareas propuesta
4	3	2	0

V**EJERCICIOS RESUELTOS**

Una vez instalado debe de lucir de la siguiente forma.

**VI****EJERCICIOS PROPUESTOS**

Proponer un conjunto de datos y desarrollar los mismos pasos que el ejercicio planteado.

VII**CUESTIONARIO**

1. Investigar sobre otras las posibles aplicaciones de Deep Learning.
2. Indicar Las ventajas del uso de Deep Learning .

Mag. Christian Alain Revilla Arroyo
Mag. Richard Smith Escobedo Quispe

VIII

BIBLIOGRAFÍA

- [1] K. Jolly. Machine Learning with scikit-learn Quick Start Guide: Classification, regression, and clustering techniques in Python. Packt Publishing Press, 2018.