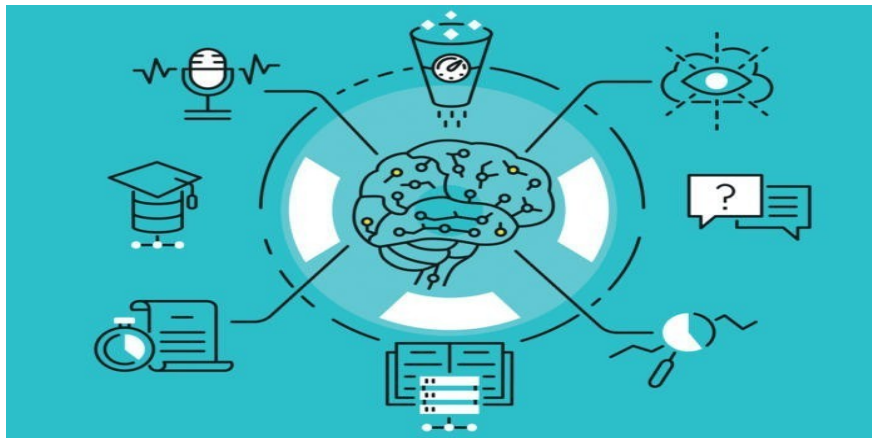




UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ciencias y Sistemas
Ingeniería de Sistemas
Inteligencia Artificial



Análisis de regresión lineal simple para una panadería.
(Documento informativo)

Integrantes:

- Br. Josué David Reyes Molina (2016-0749U)
- Br. Marcos Ricardo Millón Díaz (2015-U)
- Br. Nestor Eduardo Rivas Oviedo (2016-U)

Docente:

- Msc. Ing. Walger Herrera Treminio

Jueves 13 de febrero de 2020

Introducción

El presente documento tiene como objetivo informar acerca de la conformación y funcionalidad de lo que podría ser un modulo de análisis de datos en un sistema de información para una panadería.

La aplicación se desarrolló tomando como referencia un sistema informático que este mismo equipo de trabajo con la colaboración de Oscar Pereira y Luarín Silva ya había desarrollado para la asignatura de Producción 3, este sistema consistía en una aplicación en la que se gestionaban los inventarios de productos y de materias primas de una panadería, el cual además establecía los puntos de re-orden de los insumos.

Basándonos en la idea de mejorar la gestión administrativa de una panadería, se ha decidido desarrollar esta aplicación que en síntesis es una herramienta de análisis de datos que mediante el uso de técnicas estadísticas implementadas en algoritmos supervisados de Machine Learning realiza predicciones.

Para la implementación de estos algoritmo y por reglas que ya se habían establecido con el docente se ha utilizado el lenguaje de programación Python en su versión 3 ([Python3](#)), además de una serie de librerías que se detallan a continuación:

Prince : librería para hacer análisis factoría, que además incluye análisis de componentes principales.

Scikit-learn : Proporciona efectivas herramientas fáciles de implementar para realizar análisis de predicción de datos.

Pandas : es una herramienta fácil de análisis y manipulación de datos, rápida, potente flexible y fácil de usar.

Matplotlib: Es una biblioteca de trazado 2D que deja al alcance una variedad de formatos impresos de gráficos iterativos.

Los datos fueron recuperados de la Internet de una panadería llamada BreadBasket quienes han proporcionado información acerca de sus actividades de los año 2016 y 2017 para realizar sobre estos análisis de minería de datos y Machine Learning. Estos datos están contenidos en un dataset llamado BreadBasket_DMS.csv y cuenta con un total de 21,294 registros. Estos datos son los que en nuestra tarea de curso hemos analizado.

Además se han migrado los datos de este dataset a una base de datos relacional en el gestor de base de datos mysql, esto con el fin de poder manipular los datos de la mejor manera posible, facilitando las consultas y la programación de la aplicación.

Todos estos elementos se unen para poder consolidar una aplicación capaz de realizar una regresión lineal simple que servirá para poder realizar pronósticos de venta, para una determinada semana. Teniendo como variable dependiente las ventas en unidades y como variable independiente la semana.

Ademas como extra se han agregado una serie de ejemplos de la implementación de algoritmos de clasificación, y regresión utilizando redes neuronales. También se proporcionan una serie de herramientas para la visualización de datos como lo son las tablas, un área para realizar consultas a la base de datos y etc.

Este documento se divide en 3 secciones, la primera parte muestra los elementos del modelo de análisis de la aplicación, el escenario de aplicación, algoritmo a implementar entre otros. La segunda parte se describe una serie de pasos para hacer andar la aplicación (guía de instalación) en un ordenador tanto con sistema operativo Windows, como Linux y por ultimo una pequeña guía de usuario para inducir a las personas interesadas en su funcionamiento.

Modelado de la aplicación

1. Analizar un escenario donde se pueda implementar un algoritmo de machine learning

Nuestro proyecto permitirá realizar un pronóstico del nivel en el inventario de materiales tomando en cuenta las ventas realizadas en una determinada semana.

2. Establecer el algoritmo a utilizar (aprendizaje supervisado y no supervisado)

El algoritmo que se utilizara es el algoritmo de aprendizaje supervisado basado en un modelo de regresión, el cual consiste en predecir valores para una variable dependiente mediante una o varias variables independiente.

El modelo de regresión que se utilizara, se estará determinado por aquel que mejor se ajuste a los datos almacenado en la base de datos.

3. Requerimientos

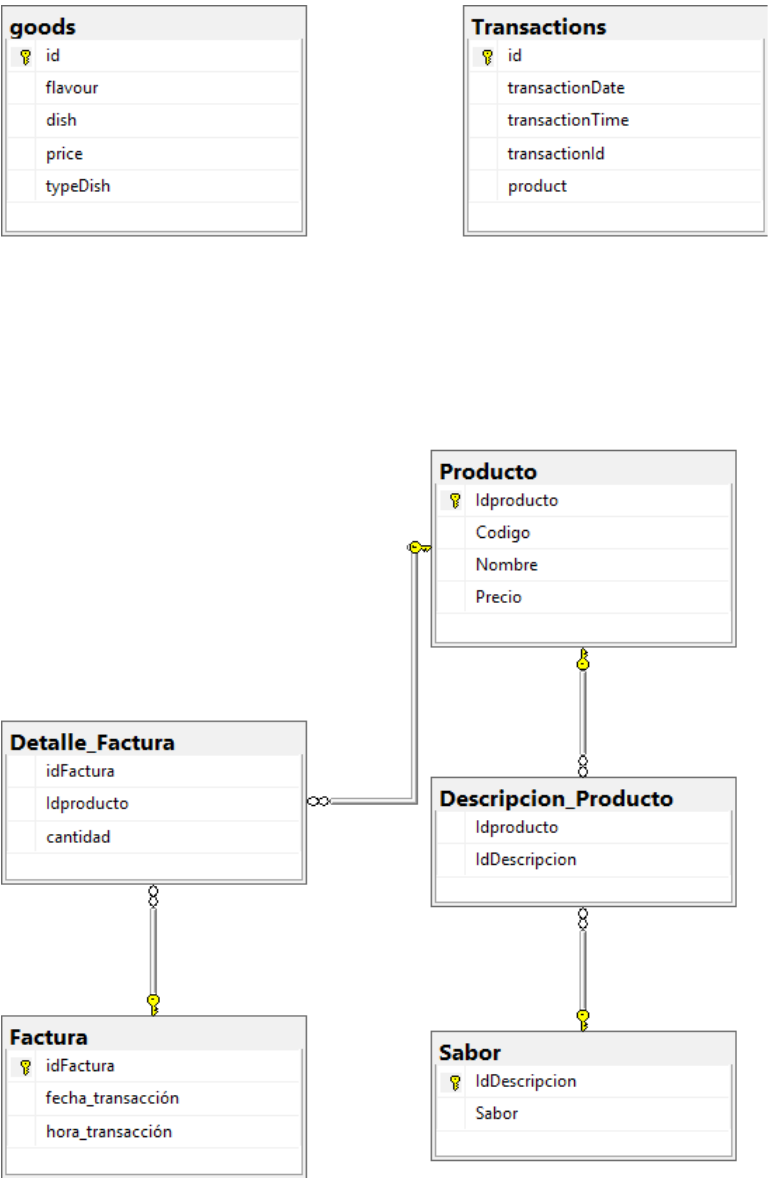
Funcionales

- Permitirá la visualización de gráficos que muestren de manera detallada el resultado de los datos analizados.
- Validación de campos que el usuario debe ingresar.
- Mostrar en una grafica la curva que mejor se ajuste a los datos que se pretende analizar.
- El sistema mostrara los detalles de venta mediante una tabla.

No funcionales:

- Facilidad de autenticación en el sistema.
- El sistema debe poseer interactividad entre los usuarios.
- Rápido tiempo de repuesta.
- Se debe mandar un mensaje de alerta cuando no se encuentre la fuente de datos.
- Los datos modificados en la base de datos se deben ver reflejados en la interfaz gráfica y viceversa.
- Las interfaces gráficas deben ayudar a comprender mejor el uso del sistema.

4. Modelo Relacional



El modelo relacional muestra como está estructurada la información en la base de datos. En el diagrama anterior se puede observar que la base de datos se conforma por 7 tablas de las cuales 5 están relacionadas entre sí y las tablas goods y Transactions que no tienen relación con ninguna otra. Esto es debido a que estas tablas han sido utilizadas simplemente para alojar los datos que estaban en el dataset BreadBasket_DMS.csv, estos se migraron de forma masiva a la tabla Transactions, en total se pasaron 21,294 registros. Luego en la tabla goods se encuentra los datos de los productos, como el precio, nombre etc.

Tomando en cuenta la base de datos que se había realizado en el sistema informático de la asignatura de Producción 3, y tomando en cuenta solo los elementos que nos interesa analizar en este trabajo se desarrolló la pequeña base de datos conformada por solamente 5 tablas: Producto, Sabor, Factura, Detalle_Factura, Detalle_Producto.

Posteriormente se procedió a mover los datos de las tablas Goods y Transactions a las tablas que conforman la base de datos. Las consultas sql que se hicieron para realizar esta migración son las que se muestran a continuación:

Cargar los datos a la tabla de productos

```
use Pastelería;
begin transaction
insert into Producto(Nombre,Codigo)
select
distinct t1.product,
(select top 1 (select dbo.fn_createCode(t2.id, t2.product))
from Transactions t2 where t2.product = t1.product) as Codigo from Transactions t1
commit transaction
```

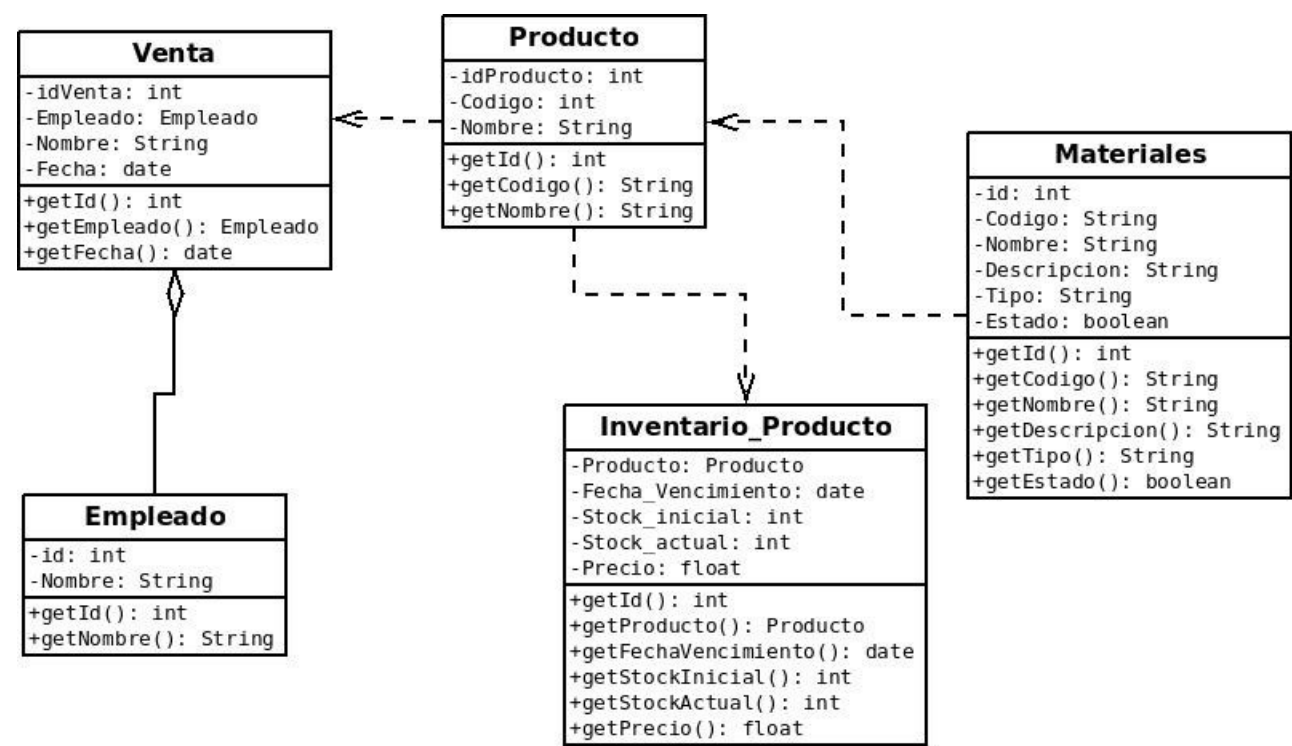
Cargar los datos a la tabla Factura

```
begin transaction
insert into Factura
select distinct t.transactionId, t.transactionDate, t.transactionTime from Transactions t order
by t.transactionDate
commit transaction
```

Cargar los datos a la tabla sabor

```
begin transaction
insert into Sabor(Sabor)
select distinct flavour from goods;
commit transaction
```

5. Diagrama de Clase

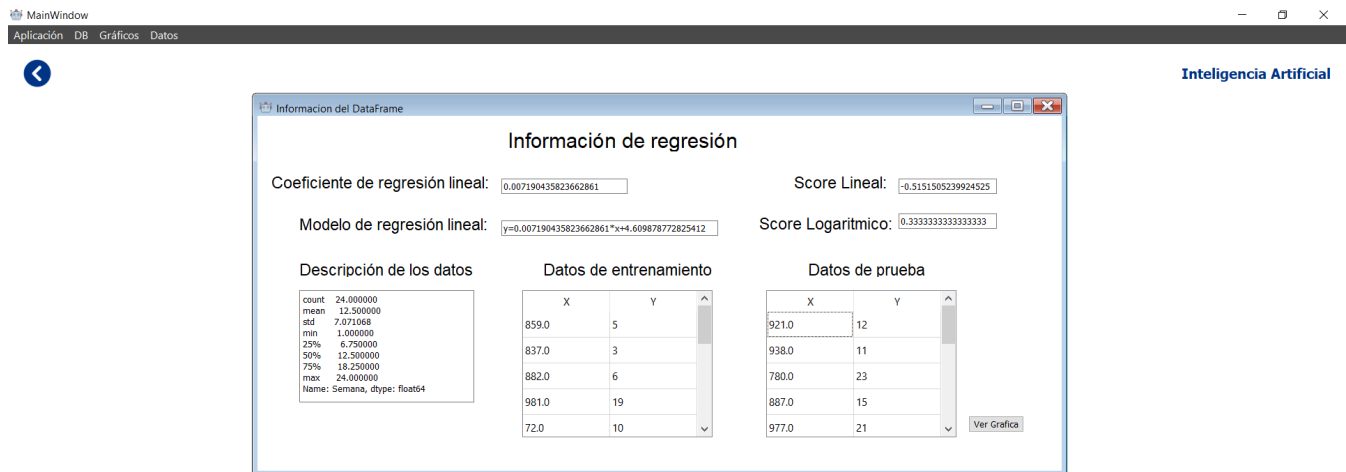


El diagrama de clases que se muestra se ha realizado ya que es un requisito solicitado por el docente, sin embargo en desarrollo de este trabajo no ha sido necesario de su implementación ya que no se esta gestionando información, solo se están haciendo consultas a las bases de datos y estos se analizan utilizando los algoritmos y luego se muestran los resultados.

Resultados

Con este proyecto se pretende brindar una herramienta que le permita a una panadería conocer sus ventas por medio de la implementación de los algoritmos de Machine Learning, específicamente el de Regresión Lineal.

Como resultados tenemos las siguientes pantallas:



Muestra la descripción de la regresión lineal, en la primera etiqueta se muestra el coeficiente de la regresión lineal, luego se muestra su score que es de -0,51 lo que indica que la relación es inversamente proporcional, luego se muestra la ecuación de regresión lineal. También se ha probado generar un modelo de regresión logarítmica que tiene un score positivo de 0,33 y podemos ver que la relación es mejor con la lineal dado que el score es mayor.

También en la primera tabla se muestra la descripción del dataset de datos que se utilizó para realizar la regresión, posterior se muestra los datos de entrenamiento de la variable dependiente (Y) y de la variable independiente (X) y también los datos que con los que se ha probado el modelo.

Dado que el modelo no se ajustó de la mejor manera a los datos, no se realizó la opción de que se generen pronósticos dado que no tendría sentido porque este modelo no serviría para pronosticar.

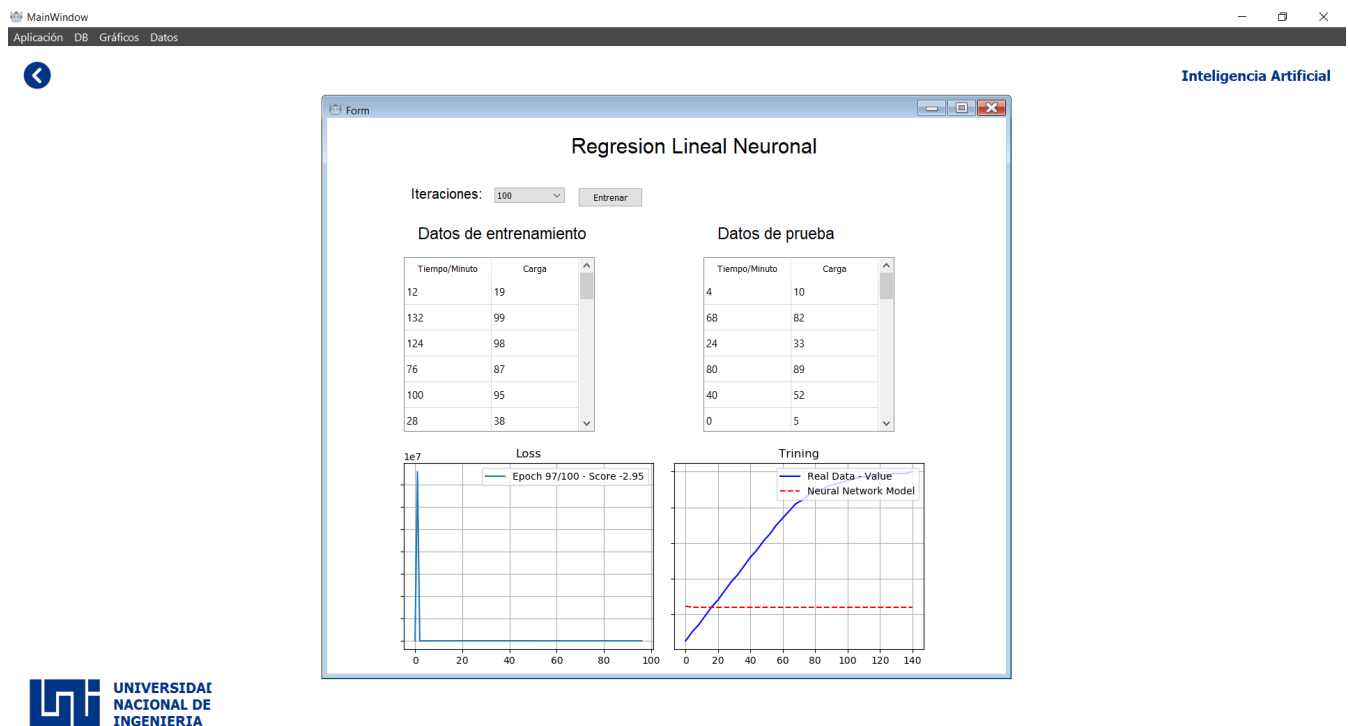
Form

Productos por semanas

Productos: * TRIAL * TRIAL ... Todos

Codigo	Producto	Semana	Cantidad
A8919	* TRIAL * TRIAL ...	1	2
F61	Fudge	1	2
T4053	Tiffin	1	1
P9	Pastry	1	13
S137	Smoothies	1	1
N27	NONE	1	41
C205	Coke	1	3

También se muestra una interfaz en la que se puede ver los datos de un producto en específico.



También en esta aplicación se han agregado opciones extras como la implementación del algoritmo de red neuronal tanto con regresión lineal como con clasificación. En la imagen anterior se muestra la red neuronal con regresión lineal.

En la imagen anterior se muestra esta regresión, algo característico de esta es que se ve en tiempo real como la grafica se ajusta al modelo de la red neuronal.

MainWindowAplicación DB Gráficos Datos

Inteligencia Artificial

Form

Red neuronal de clasificacion MLPClassifier

Separacion de datos: 0.1

Interacciones: 100

Solver: adam

Entrenar

Datos de entrenamiento

Datos de prueba

Detalle del modelo

Predicción

Plataforma:


Publicado:

Ventas globales:

Resultado:

Predecir

Matriz

 UNIVERSIDAD NACIONAL DE INGENIERIA

En la imagen anterior se muestra la red neuronal con clasificación. Este realiza pronósticos para determinar el tipo de genero del videojuego.

Instalación de la aplicación

Pasos para montar la base de datos SQL Server

1.- Ejecutar el query Panadería.sql

Este query crea la base de datos en SQL Server, en total se crean 7 tablas de las cuales 5 corresponden a la base de datos final y las otras son tablas que se utilizan para almacenar los datos temporalmente durante la migración de estos mismos a sus respectivas tablas de forma estructurada.

Además se crea un procedimiento que servirá para realizar las inserciones masivas de un dataset a la base de datos y una función que auto genera el código del producto la cual es utilizada más adelante en la inserción de estos productos.

2.- Entra a ejecutar el archivo main.py que se encuentra en el proyecto llamado migration-dataframe-sqlserver. Para esto es necesario utilizar la librería pyodbc de sqlserver y el de pandas.

Este pequeño programa inserta los datos de un dataset a la base de datos, específicamente a la tabla llamada Transactions. Hay que tener cuidado de cambiar los datos de su gestor, como el nombre del servidor, usuario y contraseña.

3.- Ejecuta el query EB-build-goods.sql

En este query se encuentran los datos del precio de los productos y estos son ingresados en la tabla goods.

4.- Ejecuta el query Datos.sql

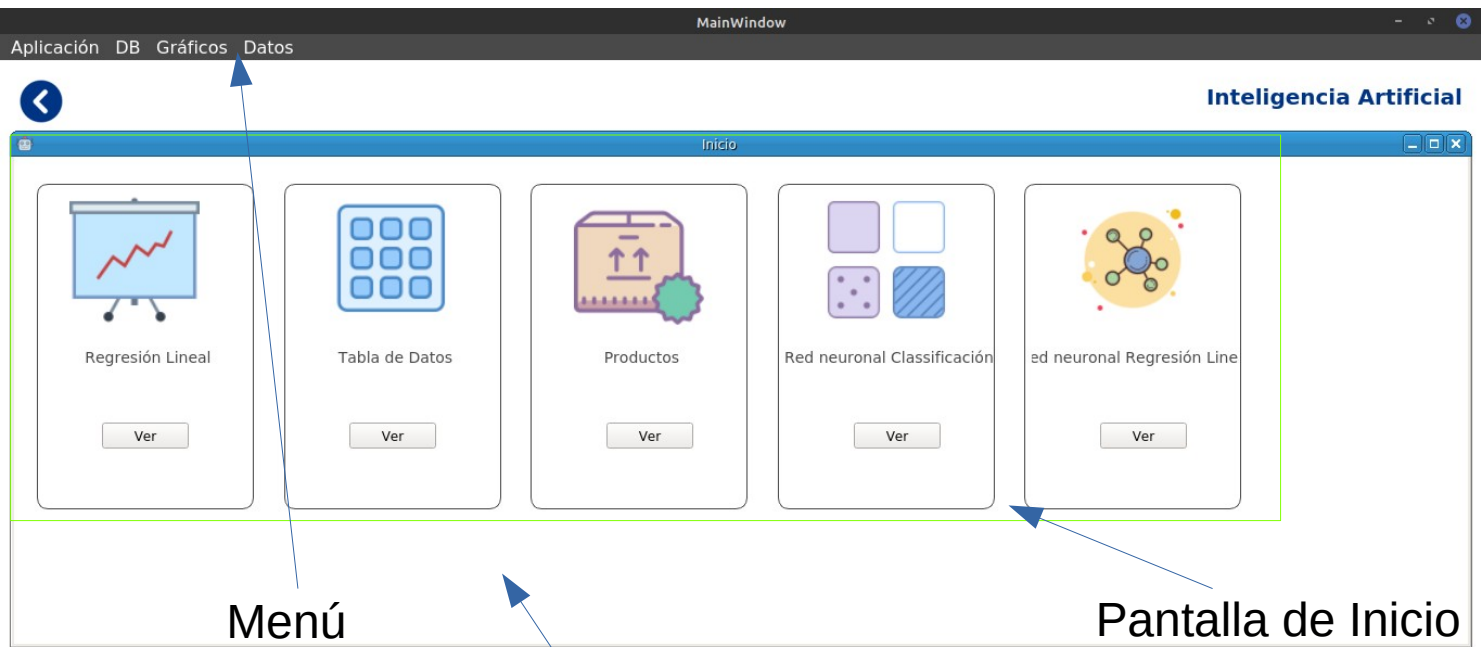
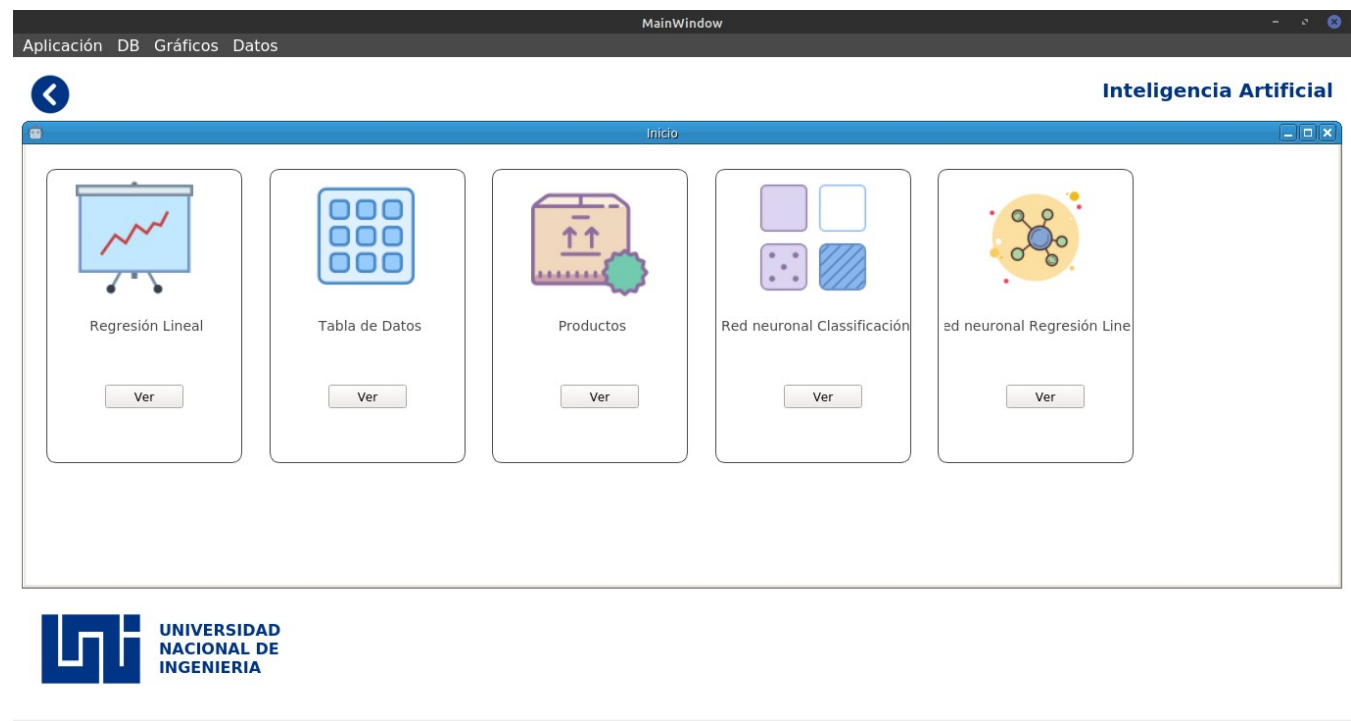
En este query se encuentran las migraciones de las tablas temporales que se llenaron por la inserción masiva del dataset a la base de datos y se reinsertan en tablas relacionales de la base de datos final.

5.- Finalmente se debe ejecutar el query Cursores.sql

Finalmente se crean cursores que insertan y actualizan datos y tablas en la base de datos, dejándola lista para su respectivo funcionamiento.

Guía de usuario

Pantalla principal





Opción Aplicación: Muestra las opciones de

Inicio → Abre la pantalla de inicio en el escritorio

Acerca → Descripción de los integrantes de este grupo

Documentación → Abre la documentación de la app

Cerrar → Cierra la aplicación



Opción DB: Cuenta con las opciones

Consultas → Para realizar consultas Sql a la DB

Diagrama MR → Muestra el diagrama de base de datos



Aplicación DB Gráficos Datos

MainWindow

Inteligencia Artificial

Frame

SQL

Consultas a la base de datos

```
SELECT week(f.fecha_transacción), sum(df.cantidad)
from detalle_factura df inner join producto p
on df.idproducto = p.idproducto
inner join factura f on df.idfactura = f.idfactura
group by week(f.fecha_transacción)
```

Resultado

Se escribe la instrucción sql

Ejecuta la instrucción

Muestra el no de filas retornadas

Tabla retornada

Aplicación DB Gráficos Datos

MainWindow

Inteligencia Artificial

Create button

Diagrama modelo relacional

goods

Transactions

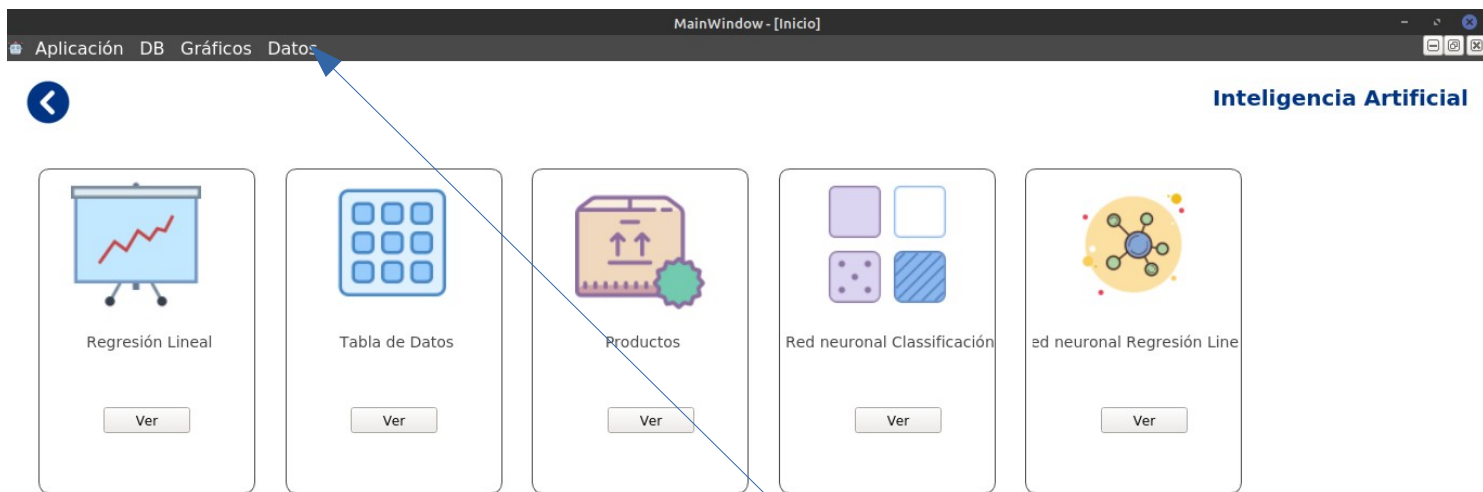
Producto

Detalle_Factura

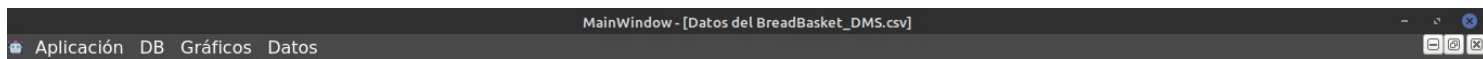
Factura

Sabor

Descripcion_Producto

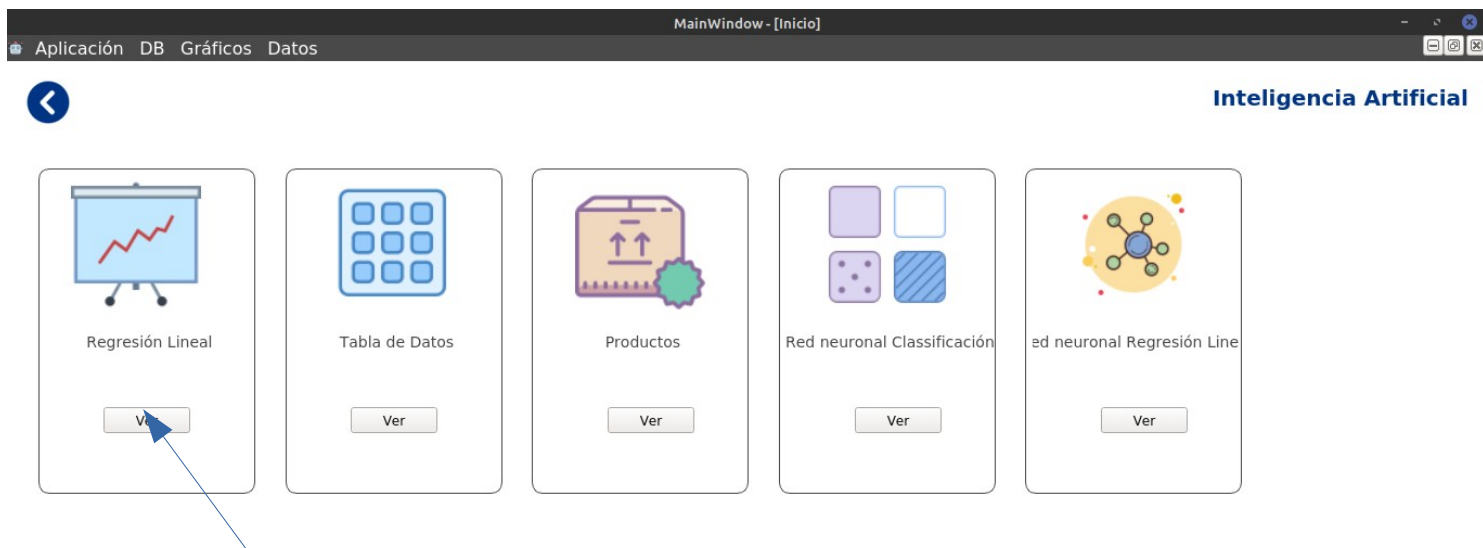


Opción datos:
Muestra el origen de los
datos, Dataset original



Dataset Original

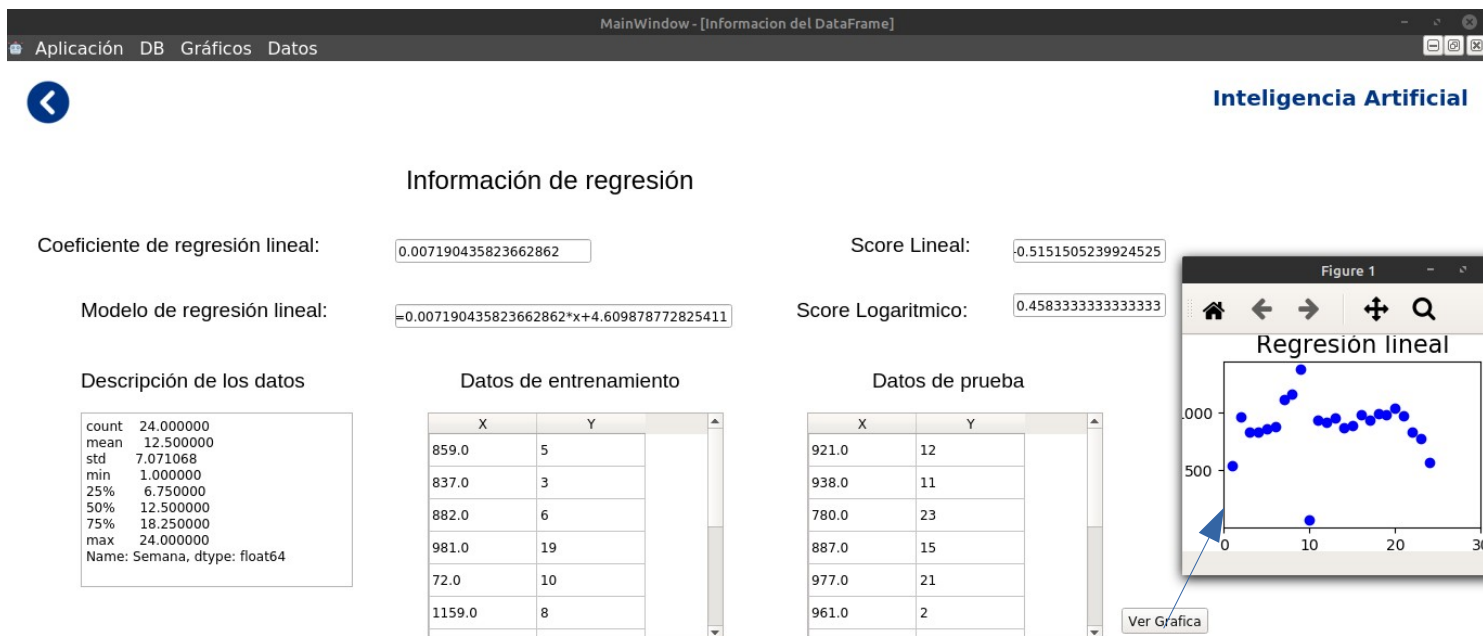
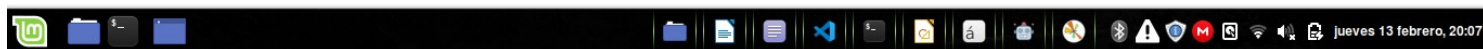
	Date	Time	Transaction	Item
0	Date	Time	Transaction	Item
1	2016-10-30	09:58:11	1	Bread
2	2016-10-30	10:05:34	2	Scandinavian
3	2016-10-30	10:05:34	2	Scandinavian
4	2016-10-30	10:07:57	3	Hot chocolate
5	2016-10-30	10:07:57	3	Jam
6	2016-10-30	10:07:57	3	Cookies
7	2016-10-30	10:08:41	4	Muffin
8	2016-10-30	10:13:03	5	Coffee
9	2016-10-30	10:13:03	5	Pastry
10	2016-10-30	10:13:03	5	Bread
11	2016-10-30	10:16:55	6	Medialuna
12	2016-10-30	10:16:55	6	Pastry



Muestra la descripción y gráficos de la regresión lineal



Y logarítmica de los datos de la panadería.



UNIVERSIDAD
NACIONAL DE
INGENIERIA

Muestra la grafica de dispersión



Muestra solo los productos
Y permite filtrar entre
estos.



Selecciona el Producto

Tabla de Productos

Form

Productos por semanas

Productos: * TRIAL * TRI... Todos

Codigo	Producto	Semana	Cantidad
A8919	* TRIAL * TRI...	1	2
F61	Fudge	1	2
T4053	Tiffin	1	1
P9	Pastry	1	13
S137	Smoothies	1	1
N27	NONE	1	41
C205	Coke	1	3
F325	Focaccia	1	2
M7	Muffin	1	12





Red neuronal de clasificación que muestra la predicción

De a que clasificación pertenece un video juego.



UNIVERSIDAD
NACIONAL DE
INGENIERIA



Red neuronal de clasificacion MLPClassifier

Separacion de datos: 0.1

Interacciones: 100

Solver: adam

Entrenar

Datos de entrenamiento

Plataforma	Autor	Ventas	Genero
4	326	0.31	Action
19	264	0.04	Adventure
26	127	0.53	Adventure
4	297	0.08	Role-Playing
26	495	0.28	Sports
7	526	0.18	Platform

Datos de prueba

Plataforma	Autor	Ventas	Genero
15	5	1.28	Action
26	348	0.07	Sports
16	139	0.14	Sports
6	86	0.15	Fighting
11	348	0.41	Puzzle
4	526	0.03	Simulation



UNIVERSIDAD
NACIONAL DE
INGENIERIA



Regresión Lineal

Ver

Tabla de Datos

Ver

Productos

Ver

Red neuronal Classificación

Ver

Red neuronal Regresión Lineal

Ver

Red neuronal de regresión lineal



Regresión Lineal Neuronal

Iteraciones:

Datos de entrenamiento

Tiempo/Minuto	Carga
116	98
128	99
84	90
48	61
140	100
8	14
20	28
0	5

Datos de prueba

Tiempo/Minuto	Carga
32	42
104	96
56	70
88	92
72	84
24	33
76	87
112	97

