





## Práctica 8

## Práctica 8:

Aplicación de tareas de aprendizaje supervisado.

## CONTENIDO

1.	OBJETIVO	2
2.	INTRODUCCIÓN	2
	DESARROLLO	
	CONCLUSIÓN	
5.	REFERENCIAS	7





#### Práctica 8

#### 1. OBJETIVO

Aplicar alguna de las tareas de aprendizaje supervisado.

#### INTRODUCCIÓN

Esta práctica se desarrollará con el lenguaje de programación Python, en la cual la principal librería que se utiliza es la librería Pandas.

Las principales características de esta librería son:

- Define nuevas estructuras de datos basadas en los arrays de la librería NumPy, pero con nuevas funcionalidades.
- Permite leer y escribir fácilmente ficheros en formato CSV, Excel y bases de datos SQL.
- Permite acceder a los datos mediante índices o nombres para filas y columnas.
- Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.
- Permite trabajar con series temporales.
- Realiza todas estas operaciones de manera muy eficiente.

Esta librería se utilizará principalmente para la extracción de datos de los archivos xls. MySQL Workbench es un software creado por la empresa Sun Microsystems, esta herramienta permite modelar diagramas de Entidad-Relación para bases de datos MySQL.

Con esta herramienta se puede elaborar una representación visual de las tablas, vistas, procedimientos almacenados y claves foráneas de la base de datos. Además, es capaz de sincronizar el modelo en desarrollo con la base de datos real. Se puede realizar una ingeniería directa e ingeniería inversa para exportare e importar el esquema de una base de datos ya existente el cual haya sido guardado o hecho copia de seguridad con MySQL Administrador.

MySQL Workbench puede generar también el guion necesario para crear la base de datos que se ha dibujado en el esquema; es compatible con los modelos de base de datos de DBDesigner 4 y soporta las novedades incorporadas en MySQL 5.x

La versión utilizada en esta práctica es la 8.0.12.

#### MySQL-connector

Cuando tenemos que almacenar gran cantidad de datos y su posterior procesamiento es muy común utilizar un gestor de bases de datos.

Con Python podemos comunicarnos con un gestor de bases de datos para enviar y recuperar datos.

Existen el gestor de bases de datos que utilizaremos desde Python será el gestor de base de datos MySQL.





#### Práctica 8

#### 2. DESARROLLO

Procedimiento: Desarrolle un dataset de entrenamiento (training) y de pruebas (testing) a partir de alguna de las bases de datos: proyecto semestral, incidentes viales o precipitación pluvial (con datos del 2019-2021).

Elija un problema de clasificación o predicción aplicado a alguna de las bases de datos mencionadas, indique en un párrafo de texto el problema elegido y que desea obtener.

 Se pronostica el número de delitos por mes y día de la semana en la delegación Gustavo A Madero.

Diseñe dos tablas de hechos, o datasets, para la fase de entrenamiento y de pruebas. Documente y explique porque eligió las dimensiones o columnas y el nivel de granularidad de datos.

idCat DiaSem_hechos Mes_hecho	os NoDelitos
-------------------------------	--------------

- idCat: Es la categoría de los delitos cometidos, este dato es un código numérico y se agregara un catálogo externo para identificar de que tipo se trata.
- DiaSem\_hechos: Es el día de la semana indicado en número, siendo 1 el día lunes y 7 el día domingo.
- Mes\_hechos: Es el mes en el que ocurrieron los hechos, esta agrupado junto con el día de la semana.
- No.Delitos: Es un conteo de los delitos que ocurrieron agrupando los campos anteriores.

Esta agrupación se realizó mediante la extracción de datos en un cubo de datos generado por la práctica anterior, el cubo agrupa los elementos de hechos, tiempo y locación.

Utilice el código adjunto con nombre "svmPredictorBetaParaModificarParaClasificador.zip" el cual consiste en un ejemplo en python para predecir valores usando SVM. También este código puede ser adaptado para emplearse en un problema de clasificación (ver las clases en video relacionadas). NO USE LOS DATASETS DE ENTRENAMIENTO QUE ESTÁN INCLUIDOS EN EL CÓDIGO FUENTE. Se recomienda usar Anaconda: <a href="https://www.anaconda.com/products/individual">https://www.anaconda.com/products/individual</a>

Documento las modificaciones del código en el reporte, junto con su respectiva explicación

No se realizaron grandes modificaciones en el código proporcionado, solo se cambiaron los nombres de los archivos y los nombres de las variables que se utilizaron.



Práctica 8



```
dfTraining = pd.read_csv("Datos_Entrenamiento.csv")
dfTesting = pd.read_csv("Datos_Prueba.csv")

X_train = dfTraining[["idCat","DiaSem_hechos","Mes_hechos"]]
y_train=dfTraining.NoDelitos

X_testing = dfTesting[["idCat","DiaSem_hechos","Mes_hechos"]]
y_testing = dfTesting.NoDelitos
```

Figura 1. Cambios en el código proporcionado.

Documente las pruebas realizadas y los desempeños obtenidos durante las pruebas. Debe realizar al menos 10 a 50 pruebas con datos diferentes (i.e. que no estén involucrados en el proceso de aprendizaje y de pruebas (80% training, 19% testing, 1% pruebas con datos que NO fueron utilizados en training ni en testing )) que se puedan comprobar, si no es así justifique su respuesta.

- Se obtuvieron un total de 881 datos después del proceso de preparación para los mismos, los cuales se guardaron en un CSV.
- Después se separaron en 3 grupos de datos (training, testing, restante) con su correspondiente porcentaje de datos proporcionado, siento un total de 704 elementos para training, 168 elementos para testing y 9 elementos restantes (no utlizados en training ni en testing) cada uno guardado en un archivo CSV distinto cada uno.
- Se utilizó el programa proporcionado para realizar la fase de training y testing, la cual funciono de manera correcta y sin errores y nos entregó dos archivos con los valores predichos Normal SVM y SVM Radial bases.
- Se agregaron los archivos salientes de la prueba de training y testing al envio.





## Práctica 8

	А	В	С	D	E		
1	predictedVa	RealValue	ErrorPercent				
2	'2.442433851	'2'	'22.121692580	0231755'			
3	'2.943253584	'2'	'47.162679210	051029'			
4	'9.188993555	'8'	'14.86241943	9968066'			
5	'4.078133644	'4'	1.953341113	597018'			
6	'6.159136892	'13'	'52.62202390	0077785'			
7	'5.966915392	'114'	'94.76586369	1031'			
8	1.097939230	'1'	9.793923047	824293'			
9	'2.310123233	'1'	'131.0123233	698235'			
10	'0.255410634	'7'	'96.351276649	94599'			
11	'2.703493550	'4'	'32.41266124	350894'			
12	'5.759366383	'13'	'55.69718166	7769755'			
13	'3.959363260	<b>'2'</b>	'97.96816302	874254'			
14	'4.146648173	'2'	'107.3324086	6505942'			
15	'5.874955231	'158'	'96.28167390	414455'			
16	'8.676900713	'22'	'60.55954220	966622'			
17	'9.359098920	'1'	'835.9098920	726814'			
18	'3.948875160	'2'	'97.44375803	312838'			
19	1.157491562	'1'	'15.74915622	1779614'			
20	'2.336852702	'1'	133.6852702	4387805'			
21	'0.781997673	'2'	'60.90011630	234549'			
22	'0.338685149	<mark>'6'</mark>	'94.35524750	30212'			
23	'2.582355695	'1'	'158.2355695	7530158'			
24	'3.904304517	'4'	'2.392387070	3475014'			
25	'5.388446933	'15'	'64.077020444	406032'			
26	'5.827095686	'142'	'95.89641148	828801'			
27	'8.102811077	'23'	'64.77038662	132736'			
28	'3.778780815	'1'	'277.87808150	0206746'			
29	'2.605201845	'1'	'160.5201845	8685095'			
30	1.406749784		'40.67497846	7401246'			
outRealTesting1							

Figura 2. Archivo de salida del programa.





#### Práctica 8

Prepare sus conclusiones indicando las lecciones aprendidas en las pruebas y con el diseño del dataset propuesto, agregue un análisis exploratorio (gráficas) del dataset del conjunto de entrenamiento (training dataset) y el conjunto de pruebas (testing dataset).

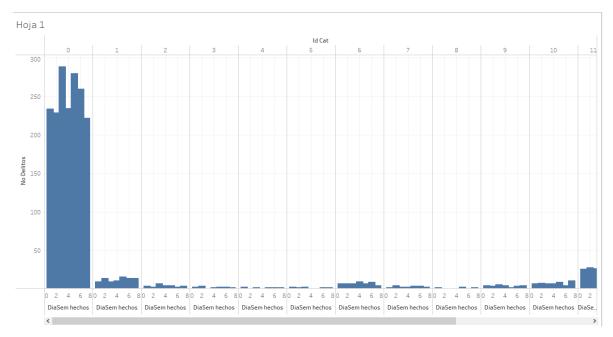


Figura 3. Grafica de Datos\_Entrenamiento

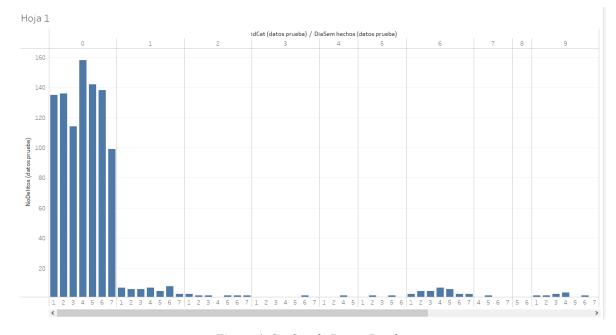


Figura 4. Grafica de Datos\_Prueba





#### Práctica 8

	Α	В	С	D	E	F	G	
1	idCAT	categoria_delito						
2	0	DELITO DE BA						
3	1	HECHO NO D	ELICTIVO					
4	2	HOMICIDIO I	DOLOSO					
5	3	LESIONES DO	LESIONES DOLOSAS POR DISPARO DE ARMA DE FUEGO					
6	4	ROBO A CAS	ROBO A CASA HABITACIÃf"N CON VIOLENCIA					
7	5	ROBO A CUE	NTAHABIENT	E SALIENDO D	EL CAJERO CO	ON VIOLENCIA	4	
8	6	ROBO A NEG	OCIO CON VI	OLENCIA				
9	7	ROBO A PAS	ROBO A PASAJERO A BORDO DE MICROBUS CON Y SIN VIOLENCIA					
10	8	ROBO A PASAJERO A BORDO DE TAXI CON VIOLENCIA						
11	9	ROBO A PAS	ROBO A PASAJERO A BORDO DEL METRO CON Y SIN VIOLENCIA					
12	10	ROBO A REPARTIDOR CON Y SIN VIOLENCIA						
13	11	ROBO A TRANSEUNTE EN VăÂA PăšBLICA CON Y SIN VIOLENCIA						
14	12	ROBO A TRA	NSPORTISTA (	CON Y SIN VIO	DLENCIA			
15	13	ROBO DE VEHÃFÂCULO CON Y SIN VIOLENCIA						
16	14	SECUESTRO						
17	15	VIOLACIÃfâ€	œΝ					

Figura 5. Catálogo de tipo de delitos

En las gráficas podemos ver que claramente en la alcaldía de Gustavo A Madero incurre en gran medida el tipo de delito 0 el cual corresponde a un delito de bajo impacto en esta delegación.

#### 3. CONCLUSIÓN

"En esta práctica pudimos utilizar los conocimientos que adquirimos en prácticas pasadas para utilizarse en conjunto para poder realizar una prueba de predicción de, en mi caso, un número de delitos que incurren en la delegación Gustavo A Madero aplicando de manera práctica lo visto en clase acerca de tareas de aprendizaje supervisado, lo cual puede ayudarnos a analizar el comportamiento de sucesos que incurren de manera recurrente con los datos que obtenemos de ciertas circunstancias".

#### 4. REFERENCIAS

- [1]. Alberca, A. S. (2020, 4 octubre). La librería Pandas. Aprende con Alf. <a href="https://aprendeconalf.es/docencia/python/manual/pandas/">https://aprendeconalf.es/docencia/python/manual/pandas/</a>
- [2]. Pública, A. D. D. I. (s. f.). Portal de Datos Abiertos de la CDMX. Portal de datos Abiertos de la CDMX. Recuperado 23 de mayo de 2021, de https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx
- [3]. Nelson Aranibar, Monografias.com. (s. f.). MySQL WorkBench. Monografias.com. Recuperado 3 de marzo de 2021, de <a href="https://www.monografias.com/trabajos88/mysql-workbench/mysql-workbench.shtml">https://www.monografias.com/trabajos88/mysql-workbench/mysql-workbench.shtml</a>