

Luis A. Muñoz M.

Maestría de Analítica de Datos

Facilitador: Juan Castillo, PhD

Panamá, 7 de septiembre de 2024

Modelos predictivos para la presentación de denuncias criminales

Proyecto final para la asignatura de Modelos Predictivos





Contenido



Antecedentes



Procedimiento
utilizado



Resultados



Conclusiones

Antecedentes



Denuncia criminal



PGN
PROCURADURÍA
GENERAL
DE LA NACIÓN

- La denuncia criminal o noticia criminal es la notificación que se hace ante el Ministerio Público de Panamá de la posible comisión de un delito que esté definido en el Código Penal.
- En términos generales, la presentación de la denuncia debe cumplir con los siguientes requisitos:
 - La persona que presenta la denuncia debe ser mayor de edad.
 - La denuncia debe ser hecha por escrito.
 - La persona que presenta la denuncia debe estar en pleno uso de sus facultades mentales.
- En Panamá las denuncias pueden tener uno de tres tipos de ingreso:
 - Por denuncia: El ciudadano se presenta ante el Ministerio Público u otra entidad receptora para dar aviso.
 - De oficio: El Ministerio Público inicia el trámite sin necesidad de que un ciudadano denuncie.
 - Por querella: El ciudadano presenta un escrito redactado por un abogado al Ministerio Público, donde se acusa a una persona, se exige resolución y se aportan pruebas. Debe ser revisada y admitida por un fiscal o juez.

Procedimiento utilizado



Procedimiento utilizado






The screenshot shows the website of the Ministerio Público de Panamá, specifically the section for judicial statistics of the Sistema Penal Acusatorio. The page has a dark blue header with navigation links: Inicio, Organización, Autoridades, Noticias, Transparencia, Judicial, and Servicios. Below the header, there is a search bar and a list of categories under the heading 'ESTADISTICAS'. The categories are: Sistema Penal Acusatorio, Anticorrupción, Estafas y Fraudes, Robos, Hurto, and Lesiones Personales. The 'Sistema Penal Acusatorio' category is selected. To the right, there is a table of statistics for the year 2016. The table has three columns: FECHA, TITULO, and ARCHIVO. The data rows show various comparative and statistical reports with PDF links.

ESTADISTICAS

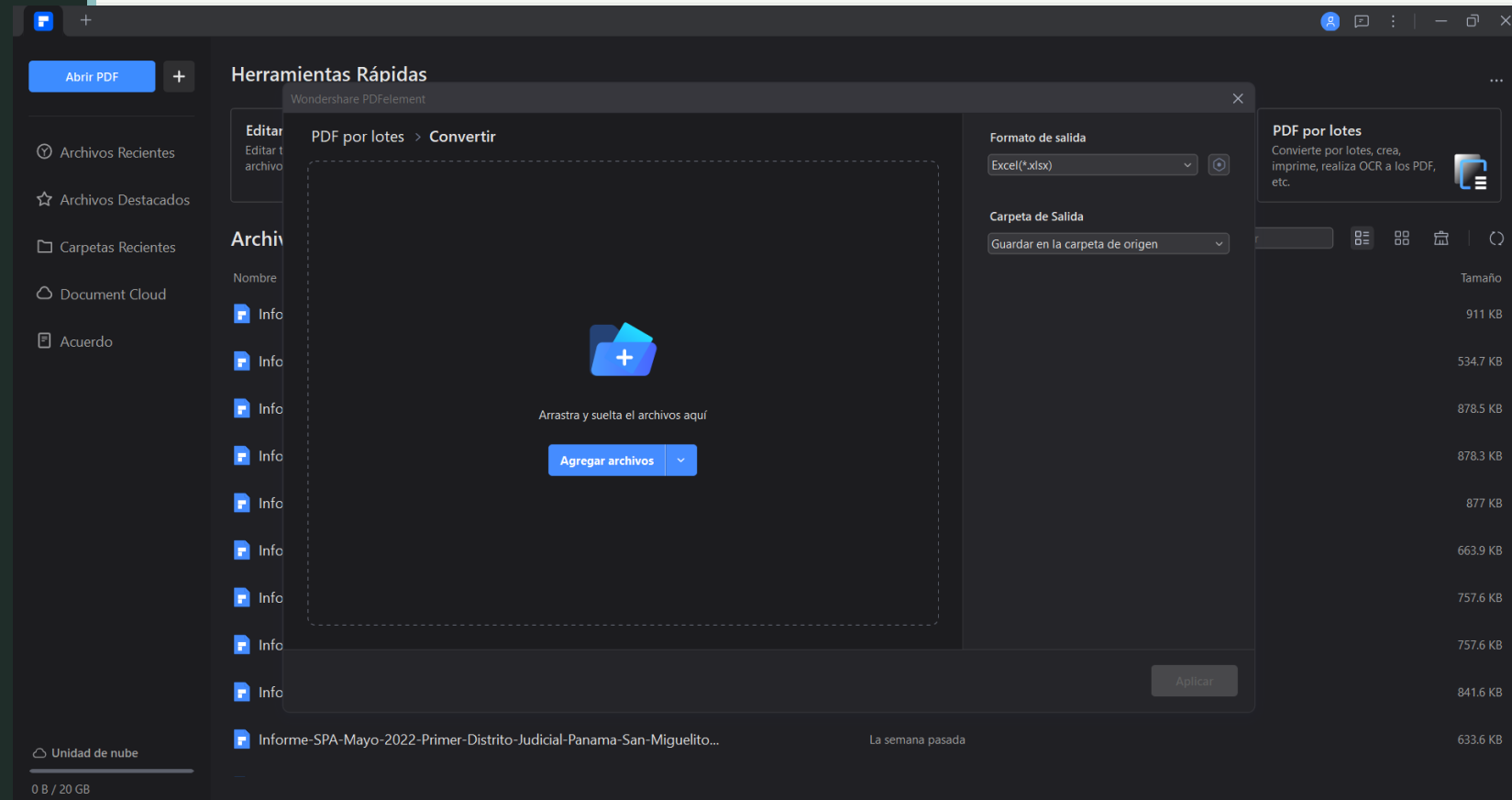
- Sistema Penal Acusatorio
- Anticorrupción
- Estafas y Fraudes
- Robos
- Hurto
- Lesiones Personales

Mostrar 15 registros Buscar: 2016

FECHA	TITULO	ARCHIVO
2016/05/05	Cuadro Comparativo Abril 2015-2016 - Segundo Distrito Judicial	
2016/05/05	Cuadro Comparativo Enero - Abril 2015-2016 - Segundo Distrito Judicial	
2016/02/05	Informe estadístico 1 al 31 de Enero 2016 – Cuarto Distrito Judicial	
2016/06/05	Informe estadístico 2 Septiembre 2015 al 31 de Mayo 2016 – Tercer Distrito Judicial	


- Se obtuvieron los datos del sitio web del Ministerio Público de Panamá, en la sección de estadísticas judiciales. La dirección es <https://ministeriopublico.gob.pa/estadisticas-judiciales/>.
- Se eligió la opción de Sistema Penal Acusatorio, en la cual reposan los informes mensuales de todas las denuncias presentadas al Ministerio Público desde la implementación del sistema (2015).

Procedimiento utilizado



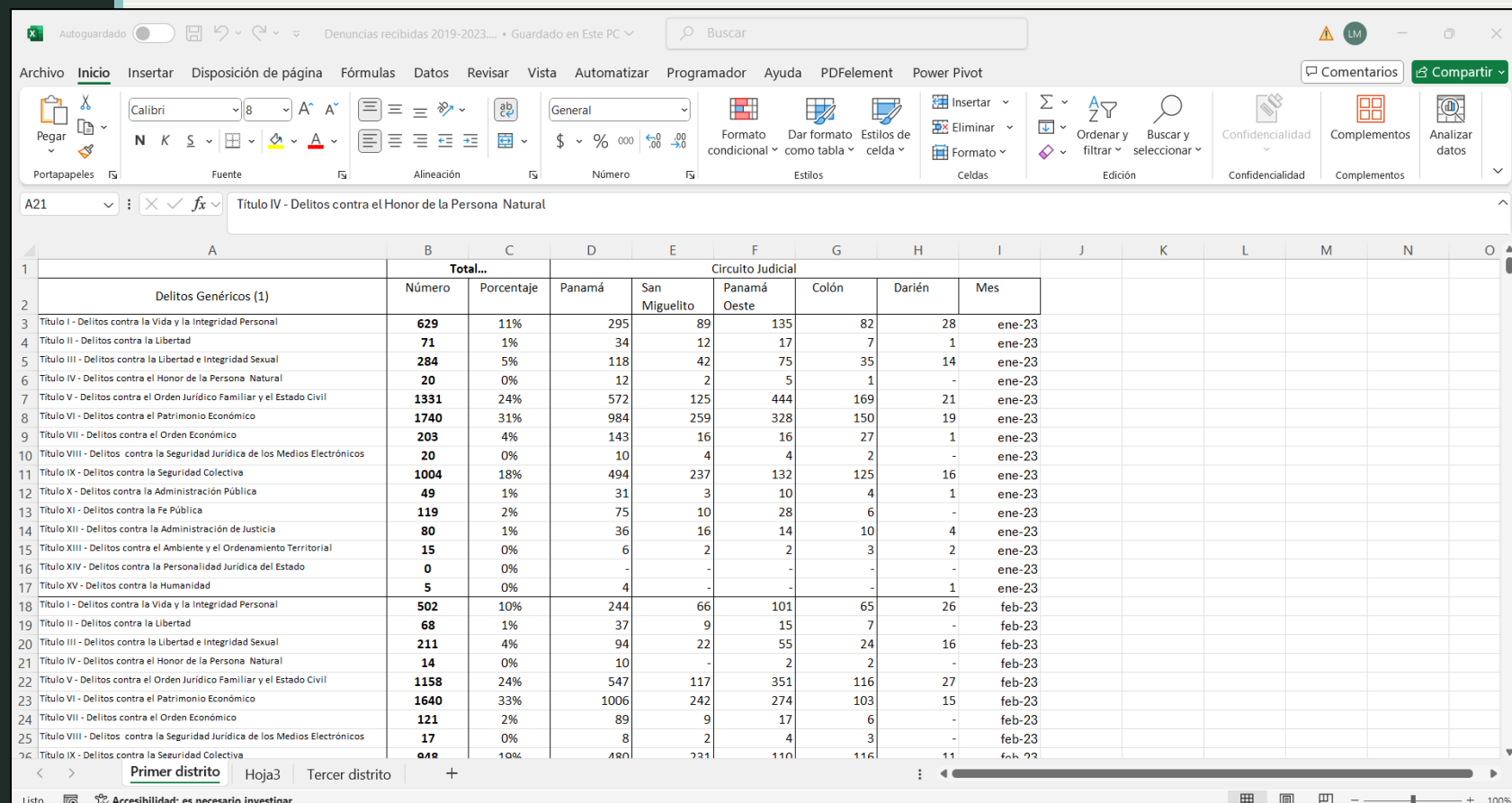
- Se seleccionaron y descargaron los archivos del primer y tercer distritos judiciales, que corresponden a los meses desde enero de 2019 hasta diciembre de 2023.
- Como los informes solamente existen en formato PDF, con el software Wondershare PDFelement se convirtieron por lotes los archivos al formato Excel.

Procedimiento utilizado

 MINISTERIO PÚBLICO PROCURADURÍA GENERAL DE LA NACIÓN				
Cuadro 2. NÚMERO DE DENUNCIAS RECEPTADAS EN EL SISTEMA PENAL ACUSATORIO DE HECHOS DELICTIVOS OCURRIDOS EN EL TERCER DISTRITO JUDICIAL DE PANAMÁ, POR CIRCUITO JUDICIAL, SEGÚN DELITOS GENÉRICOS: DEL 1 AL 30 DE SEPTIEMBRE DE 2019. (P)				
Delitos Genéricos	Total...		Circuito Judicial	
	Número	Porcentaje	Chiriquí	Bocas del Toro
TOTAL...	1001	100%	714	287
Título I - Delitos contra la Vida y la Integridad Personal	159	16%	91	68
Título II - Delitos contra la Libertad	40	4%	36	4
Título III - Delitos contra la Libertad e Integridad Sexual	127	13%	72	55
Título IV - Delitos contra el Honor de la Persona Natural	4	0%	3	1
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	232	23%	168	64
Título VI - Delitos contra el Patrimonio Económico	223	22%	187	36
Título VII - Delitos contra el Orden Económico	27	3%	25	2
Título VIII - Delitos contra la Seguridad Jurídica de los Medios Electrónicos	1	0%	1	-
Título IX - Delitos contra la Seguridad Colectiva	145	14%	103	42
Título X - Delitos contra la Administración Pública	13	1%	5	8
Título XI - Delitos contra la Fe Pública	8	1%	7	1
Título XII - Delitos contra la Administración de Justicia	17	2%	11	6
Título XIII - Delitos contra el Ambiente y el Ordenamiento Territorial	5	0%	5	-
Título XIV - Delitos contra la Personalidad Jurídica del Estado	0	0%	-	-
Título XV - Delitos contra la Humanidad	0	0%	-	-
(1) Las noticias criminales presentadas corresponden al lugar de ocurrencia del hecho.				
(P) Cifras preliminares.				
Fuente: Centro de Estadísticas, Unidad Regional de Chiriquí y Bocas del Toro. Plataforma del Sistema Penal Acusatorio.				

- Con Microsoft Excel, de cada informe mensual se copiaron los datos de la tabla de “Número de denuncias receptadas” hacia un archivo de Excel donde se consolidaron los datos mensuales.

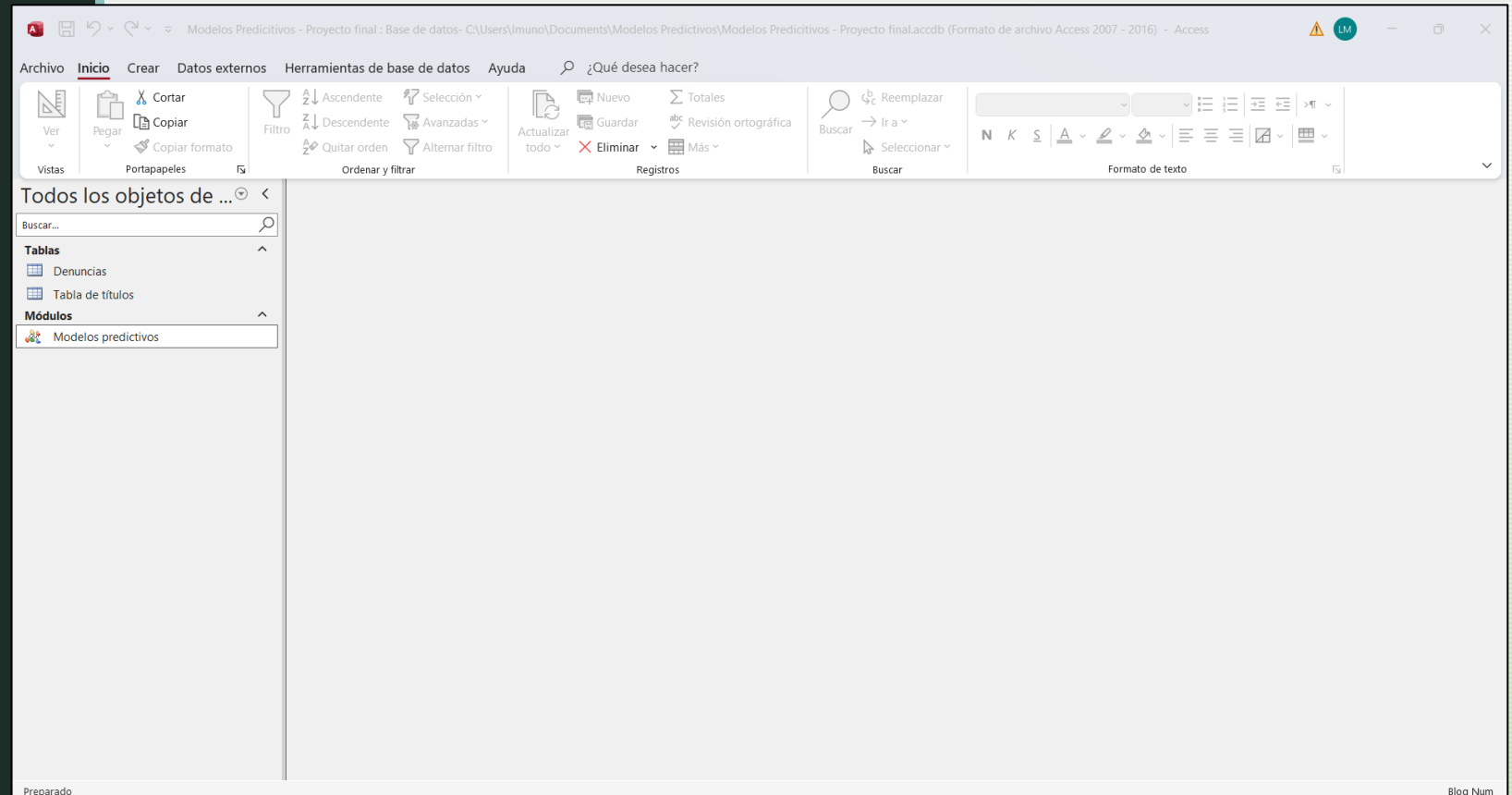
Procedimiento utilizado



Total...		Circuito Judicial						
Delitos Genéricos (1)		Número	Porcentaje	Panamá	San Miguelito	Panamá Oeste	Colón	Darién
Título I - Delitos contra la Vida y la Integridad Personal		629	11%	295	89	135	82	28
Título II - Delitos contra la Libertad		71	1%	34	12	17	7	1
Título III - Delitos contra la Libertad e Integridad Sexual		284	5%	118	42	75	35	14
Título IV - Delitos contra el Honor de la Persona Natural		20	0%	12	2	5	1	-
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil		1331	24%	572	125	444	169	21
Título VI - Delitos contra el Patrimonio Económico		1740	31%	984	259	328	150	19
Título VII - Delitos contra el Orden Económico		203	4%	143	16	16	27	1
Título VIII - Delitos contra la Seguridad Jurídica de los Medios Electrónicos		20	0%	10	4	4	2	-
Título IX - Delitos contra la Seguridad Colectiva		1004	18%	494	237	132	125	16
Título X - Delitos contra la Administración Pública		49	1%	31	3	10	4	1
Título XI - Delitos contra la Fe Pública		119	2%	75	10	28	6	-
Título XII - Delitos contra la Administración de Justicia		80	1%	36	16	14	10	4
Título XIII - Delitos contra el Ambiente y el Ordenamiento Territorial		15	0%	6	2	2	3	2
Título XIV - Delitos contra la Personalidad Jurídica del Estado		0	0%	-	-	-	-	-
Título XV - Delitos contra la Humanidad		5	0%	4	-	-	-	1
Título I - Delitos contra la Vida y la Integridad Personal		502	10%	244	66	101	65	26
Título II - Delitos contra la Libertad		68	1%	37	9	15	7	-
Título III - Delitos contra la Libertad e Integridad Sexual		211	4%	94	22	55	24	16
Título IV - Delitos contra el Honor de la Persona Natural		14	0%	10	-	2	2	-
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil		1158	24%	547	117	351	116	27
Título VI - Delitos contra el Patrimonio Económico		1640	33%	1006	242	274	103	15
Título VII - Delitos contra el Orden Económico		121	2%	89	9	17	6	-
Título VIII - Delitos contra la Seguridad Jurídica de los Medios Electrónicos		17	0%	8	2	4	3	-
Título IX - Delitos contra la Seguridad Colectiva		948	10%	480	221	110	116	11

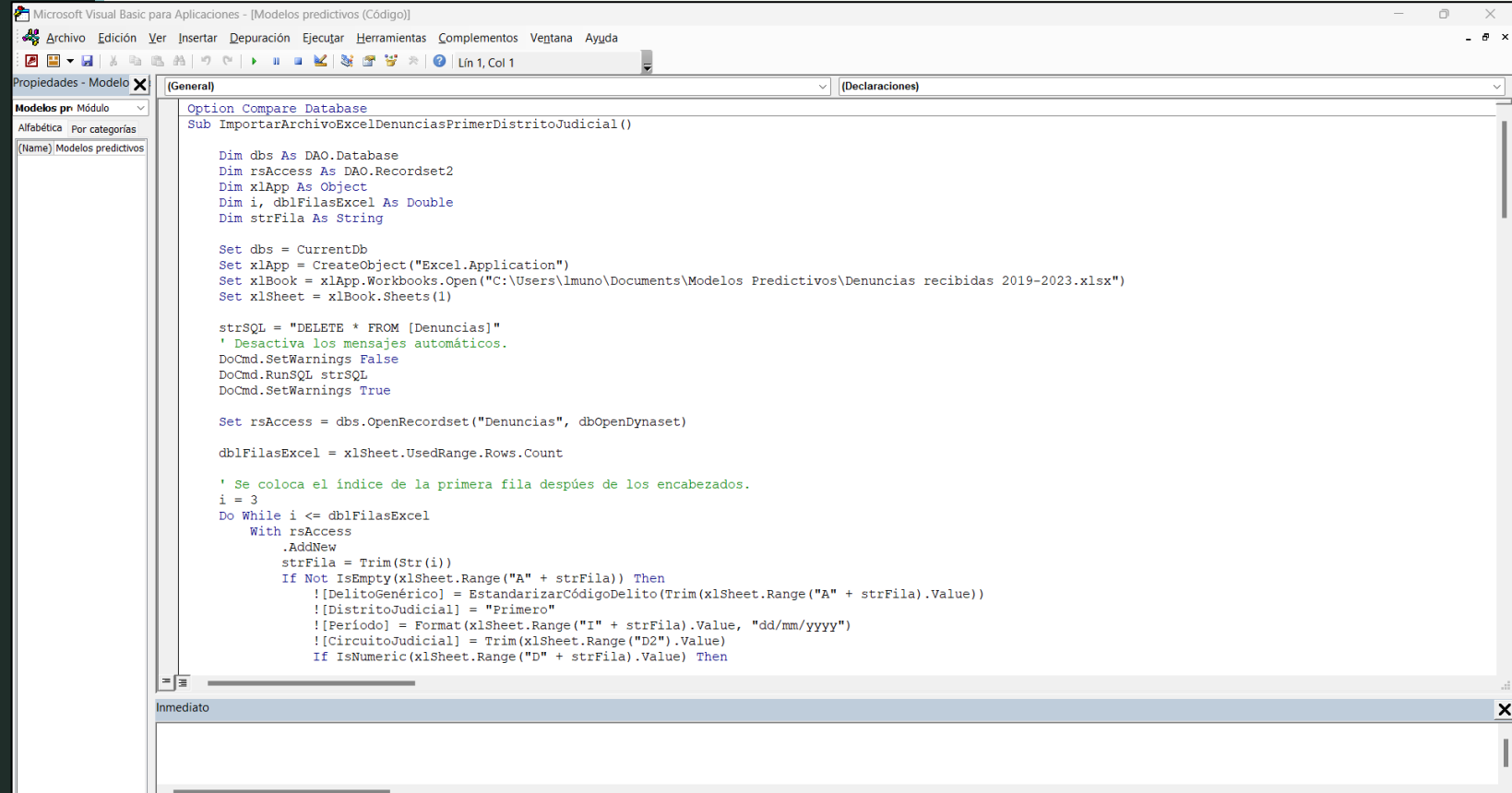
- Para cada conjunto copiado de registros se le agregó en una columna la fecha (mes y año) a la que corresponden los datos.

Procedimiento utilizado



- Debido a que los datos por circuito judicial estaban en columnas y no en filas, se creó una base de datos en Microsoft Access para normalizar los datos, es decir para que cada atributo fuera único en cada registro y aumentar la granularidad para los análisis.
- En la base de datos se creó una tabla llamada Denuncias, para almacenar los registros normalizados de denuncias por mes y por circuito judicial.

Procedimiento utilizado



The screenshot shows the Microsoft Visual Basic for Applications editor window. The title bar reads "Microsoft Visual Basic para Aplicaciones - [Modelos predictivos (Código)]". The menu bar includes Archivo, Edición, Ver, Insertar, Depuración, Ejecutar, Herramientas, Complementos, Ventana, and Ayuda. The status bar at the bottom indicates "Lín 1, Col 1".

On the left, the "Propiedades - Modelo" pane shows "Modelos por Módulo" with a list containing "(Name) Modelos predictivos".

The main editor area displays the following VBA code:

```
Option Compare Database
Sub ImportarArchivoExcelDenunciasPrimerDistritoJudicial()

    Dim dbs As DAO.Database
    Dim rsAccess As DAO.Recordset2
    Dim xlApp As Object
    Dim i, dblFilasExcel As Double
    Dim strFila As String

    Set dbs = CurrentDb
    Set xlApp = CreateObject("Excel.Application")
    Set xlBook = xlApp.Workbooks.Open("C:\Users\lmuno\Documents\Modelos Predictivos\Denuncias recibidas 2019-2023.xlsx")
    Set xlSheet = xlBook.Sheets(1)

    strSQL = "DELETE * FROM [Denuncias]"
    ' Desactiva los mensajes automáticos.
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    DoCmd.SetWarnings True

    Set rsAccess = dbs.OpenRecordset("Denuncias", dbOpenDynaset)

    dblFilasExcel = xlSheet.UsedRange.Rows.Count

    ' Se coloca el índice de la primera fila después de los encabezados.
    i = 3
    Do While i <= dblFilasExcel
        With rsAccess
            .AddNew
            strFila = Trim(Str(i))
            If Not IsEmpty(xlSheet.Range("A" + strFila)) Then
                ![DelitoGenérico] = EstandarizarCodigoDelito(Trim(xlSheet.Range("A" + strFila).Value))
                ![DistritoJudicial] = "Primerero"
                ![Período] = Format(xlSheet.Range("I" + strFila).Value, "dd/mm/yyyy")
                ![CircuitoJudicial] = Trim(xlSheet.Range("D2").Value)
                If IsNumeric(xlSheet.Range("D" + strFila).Value) Then
```

- Se creó un módulo Visual Basic para leer el archivo de Excel y separar las filas en registros por mes y por circuito judicial.
- Como los archivos del primer distrito judicial tienen más columnas que los archivos del tercer distrito judicial, en el módulo se crearon dos rutinas con el fin de leer por separado cada hoja de Excel y homologar los datos en la tabla "Denuncias".

Procedimiento utilizado

DelitoGenérico	DistrictoJudic	CircuitoJudicial	Denuncias	Periodo
Título XV - Delitos contra la Humanidad	Primero	Darién	1	01/01/2023
Título I - Delitos contra la Vida y la Integridad Personal	Primero	Panamá	244	01/02/2023
Título I - Delitos contra la Vida y la Integridad Personal	Primero	San Miguelito	66	01/02/2023
Título I - Delitos contra la Vida y la Integridad Personal	Primero	Panamá Oeste	101	01/02/2023
Título I - Delitos contra la Vida y la Integridad Personal	Primero	Colón	65	01/02/2023
Título I - Delitos contra la Vida y la Integridad Personal	Primero	Darién	26	01/02/2023
Título II - Delitos contra la Libertad	Primero	Panamá	37	01/02/2023
Título II - Delitos contra la Libertad	Primero	San Miguelito	9	01/02/2023
Título II - Delitos contra la Libertad	Primero	Panamá Oeste	15	01/02/2023
Título II - Delitos contra la Libertad	Primero	Colón	7	01/02/2023
Título II - Delitos contra la Libertad	Primero	Darién	0	01/02/2023
Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Panamá	94	01/02/2023
Título III - Delitos contra la Libertad e Integridad Sexual	Primero	San Miguelito	22	01/02/2023
Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Panamá Oeste	55	01/02/2023
Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Colón	24	01/02/2023
Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Darién	16	01/02/2023
Título IV - Delitos contra el Honor de la Persona Natural	Primero	Panamá	10	01/02/2023
Título IV - Delitos contra el Honor de la Persona Natural	Primero	San Miguelito	0	01/02/2023
Título IV - Delitos contra el Honor de la Persona Natural	Primero	Panamá Oeste	2	01/02/2023
Título IV - Delitos contra el Honor de la Persona Natural	Primero	Colón	2	01/02/2023
Título IV - Delitos contra el Honor de la Persona Natural	Primero	Darién	0	01/02/2023
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Panamá	547	01/02/2023
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	San Miguelito	117	01/02/2023
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Panamá Oeste	351	01/02/2023
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Colón	116	01/02/2023
Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Darién	27	01/02/2023
Título VI - Delitos contra el Patrimonio Económico	Primero	Panamá	1006	01/02/2023
Título VI - Delitos contra el Patrimonio Económico	Primero	San Miguelito	242	01/02/2023
Título VI - Delitos contra el Patrimonio Económico	Primero	Panamá Oeste	274	01/02/2023

- Se ejecutó el código del módulo Visual Basic .
- Se exportó la tabla “Denuncias” a un nuevo archivo de Excel. Este archivo fue la base para recopilar los datos del proyecto.

Procedimiento utilizado

Autoguardado MP Proyecto final... Guardado en Este PC

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Automatizar Programador Ayuda PDFelement Power Pivot

Comentarios Compartir

Portapapeles Pegar Fuente Alineación Número Estilos Celdas Edición

A48 Título X - Delitos contra la Administración Pública

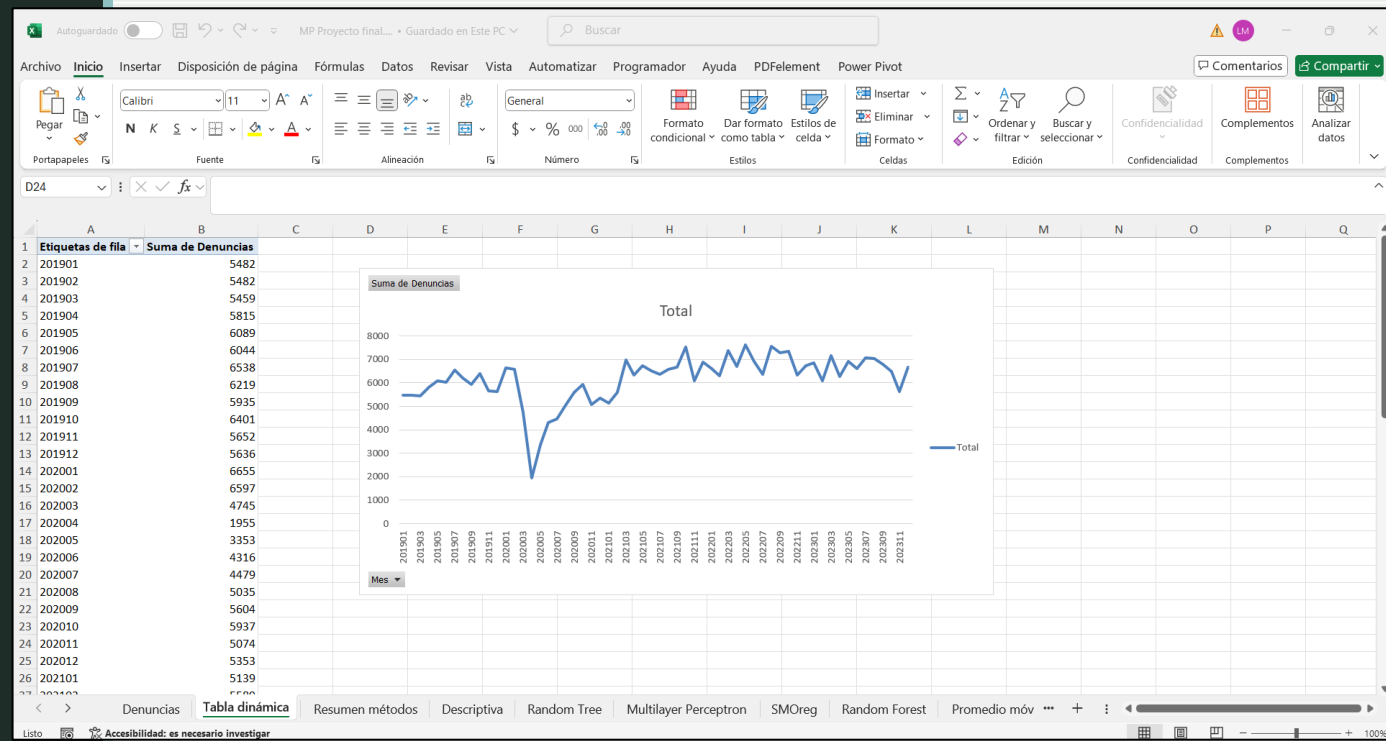
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Delito Genérico	Distrito Judicial	Circuito Judicial	Denuncias	Período	Mes								
1	Título XV - Delitos contra la Humanidad	Primero	Darién	1	01/01/2023	202301								
2	Título I - Delitos contra la Vida y la Integridad Personal	Primero	Panamá	244	01/02/2023	202302								
3	Título I - Delitos contra la Vida y la Integridad Personal	Primero	San Miguelito	66	01/02/2023	202302								
4	Título I - Delitos contra la Vida y la Integridad Personal	Primero	Panamá Oeste	101	01/02/2023	202302								
5	Título I - Delitos contra la Vida y la Integridad Personal	Primero	Colón	65	01/02/2023	202302								
6	Título I - Delitos contra la Vida y la Integridad Personal	Primero	Darién	26	01/02/2023	202302								
7	Título II - Delitos contra la Libertad	Primero	Panamá	37	01/02/2023	202302								
8	Título II - Delitos contra la Libertad	Primero	San Miguelito	9	01/02/2023	202302								
9	Título II - Delitos contra la Libertad	Primero	Panamá Oeste	15	01/02/2023	202302								
10	Título II - Delitos contra la Libertad	Primero	Colón	7	01/02/2023	202302								
11	Título II - Delitos contra la Libertad	Primero	Darién	0	01/02/2023	202302								
12	Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Panamá	94	01/02/2023	202302								
13	Título III - Delitos contra la Libertad e Integridad Sexual	Primero	San Miguelito	22	01/02/2023	202302								
14	Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Panamá Oeste	55	01/02/2023	202302								
15	Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Colón	24	01/02/2023	202302								
16	Título III - Delitos contra la Libertad e Integridad Sexual	Primero	Darién	16	01/02/2023	202302								
17	Título IV - Delitos contra el Honor de la Persona Natural	Primero	Panamá	10	01/02/2023	202302								
18	Título IV - Delitos contra el Honor de la Persona Natural	Primero	San Miguelito	0	01/02/2023	202302								
19	Título IV - Delitos contra el Honor de la Persona Natural	Primero	Panamá Oeste	2	01/02/2023	202302								
20	Título IV - Delitos contra el Honor de la Persona Natural	Primero	Colón	2	01/02/2023	202302								
21	Título IV - Delitos contra el Honor de la Persona Natural	Primero	Darién	0	01/02/2023	202302								
22	Título IV - Delitos contra el Honor de la Persona Natural	Primero	Panamá	547	01/02/2023	202302								
23	Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	San Miguelito	117	01/02/2023	202302								
24	Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Panamá Oeste	351	01/02/2023	202302								
25	Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Colón	116	01/02/2023	202302								
26	Título V - Delitos contra el Orden Jurídico Familiar y el Estado Civil	Primero	Darién	27	01/02/2023	202302								

Denuncias Tabla dinámica Resumen métodos Descriptiva Random Tree Multilayer Perceptron SMOreg Random Forest Promedio móvil

Accesibilidad: es necesario investigar

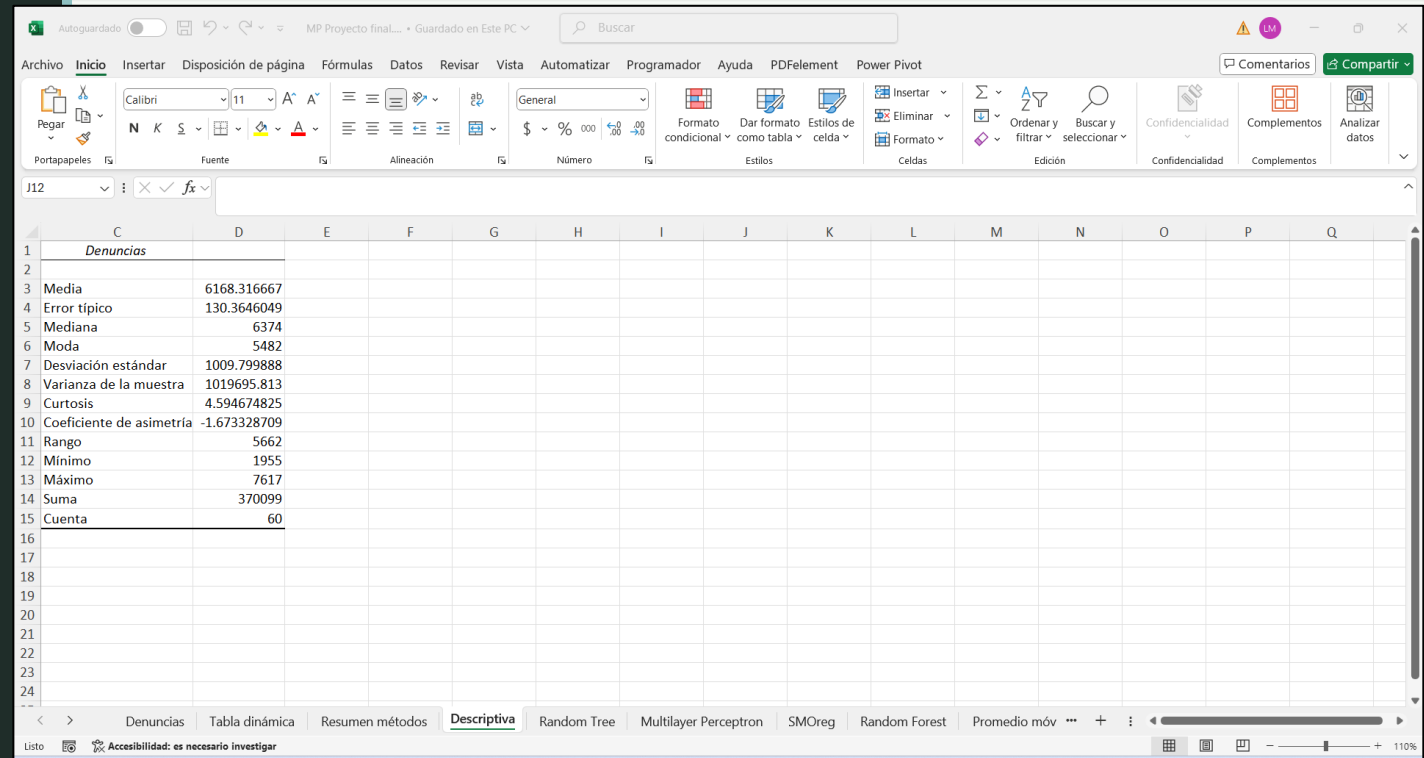
- Con el fin de evitar problemas con los formatos de fecha, a la hoja de datos de las denuncias se le agregó una columna que únicamente contiene el año y el mes de la fecha en formato numérico.

Procedimiento utilizado



- Se agregó una hoja de Excel al archivo. En ella se crearon una tabla dinámica y un gráfico dinámico, con la suma de la cantidad de denuncias agrupadas por período (año y mes). Esto permitió crear la serie de tiempo para la cantidad de denuncias presentadas.
- A partir de la tabla dinámica, se eligieron meses al azar y se cotejaron los resultados de la tabla con el archivo PDF correspondiente. Se validó que cuadraran los totales de denuncias presentadas.

Procedimiento utilizado



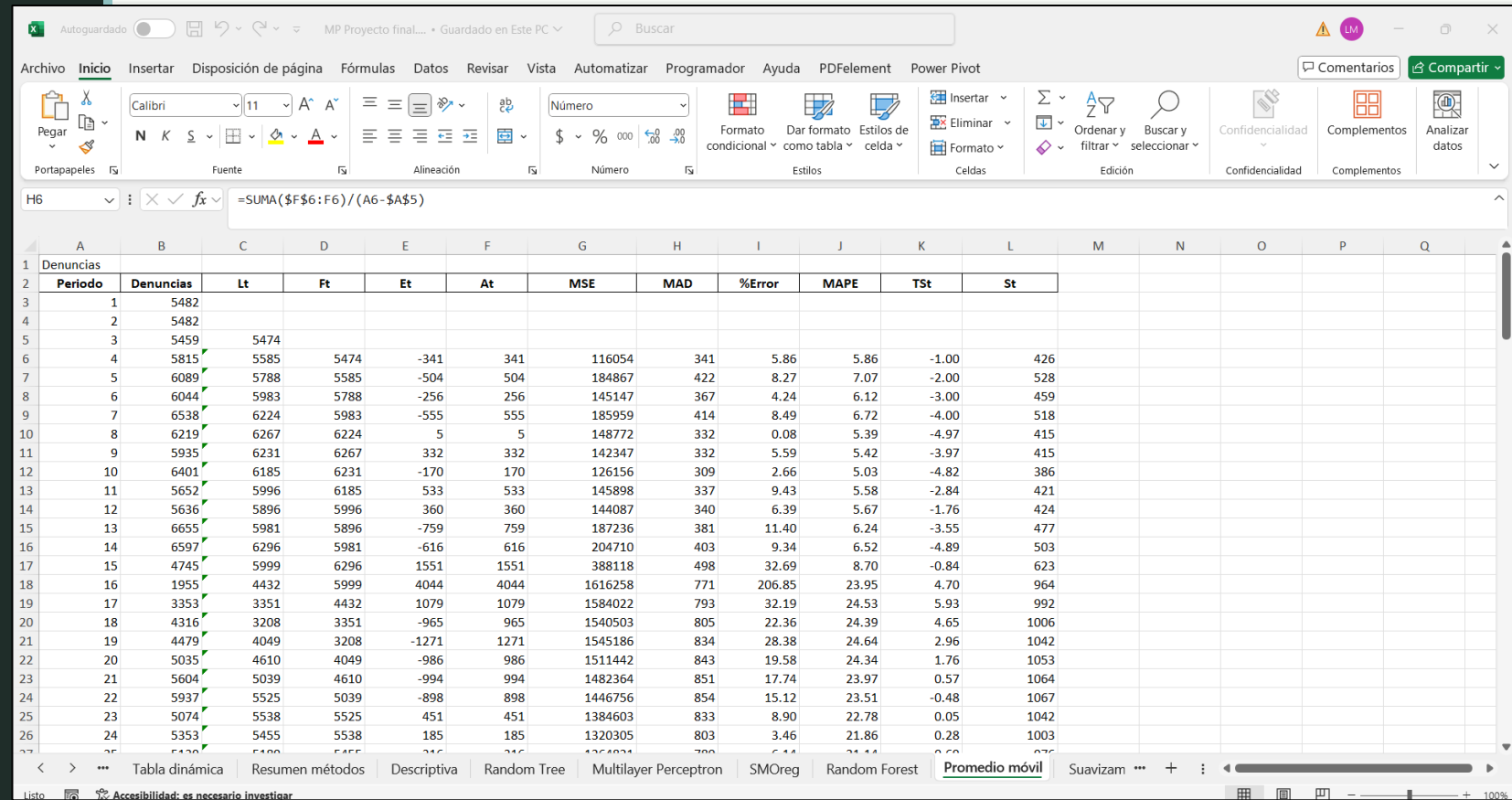
The screenshot shows the Microsoft Excel interface with a file named 'MP Proyecto final...' open. The 'Inicio' tab is active, and the ribbon shows various options like Font, Alignment, and Numbers. The spreadsheet displays a table of descriptive statistics for a dataset named 'Denuncias'.

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Denuncias														
2															
3	Media	6168.316667													
4	Error típico	130.3646049													
5	Mediana	6374													
6	Moda	5482													
7	Desviación estándar	1009.799888													
8	Varianza de la muestra	1019695.813													
9	Curtosis	4.594674825													
10	Coefficiente de asimetría	-1.673328709													
11	Rango	5662													
12	Mínimo	1955													
13	Máximo	7617													
14	Suma	370099													
15	Cuenta	60													
16															
17															
18															
19															
20															
21															
22															
23															
24															

The bottom of the window shows the 'Descriptiva' sheet selected in the 'Tablas dinámicas' group, with other sheets like 'Random Tree' and 'Multilayer Perceptron' visible.

- Se preparó una evaluación rápida de estadística descriptiva de los datos.
- Con la ayuda del gráfico dinámico se revisó si existía estacionalidad en los datos.
- Dentro del archivo de Excel se crearon hojas separadas para trabajar los pronósticos por cada método clásico y por cada algoritmo de machine learning seleccionado.

Procedimiento utilizado



The screenshot displays a Microsoft Excel spreadsheet with the following data table:

Periodo	Denuncias	Lt	Ft	Et	At	MSE	MAD	%Error	MAPE	TSst	St
1	5482										
2	5482										
3	5459	5474									
4	5815	5585	5474	-341	341	116054	341	5.86	5.86	-1.00	426
5	6089	5788	5585	-504	504	184867	422	8.27	7.07	-2.00	528
6	6044	5983	5788	-256	256	145147	367	4.24	6.12	-3.00	459
7	6538	6224	5983	-555	555	185959	414	8.49	6.72	-4.00	518
8	6219	6267	6224	5	5	148772	332	0.08	5.39	-4.97	415
9	5935	6231	6267	332	332	142347	332	5.59	5.42	-3.97	415
10	6401	6185	6231	-170	170	126156	309	2.66	5.03	-4.82	386
11	5652	5996	6185	533	533	145898	337	9.43	5.58	-2.84	421
12	5636	5896	5996	360	360	144087	340	6.39	5.67	-1.76	424
13	6655	5981	5896	-759	759	187236	381	11.40	6.24	-3.55	477
14	6597	6296	5981	-616	616	204710	403	9.34	6.52	-4.89	503
15	4745	5999	6296	1551	1551	388118	498	32.69	8.70	-0.84	623
16	1955	4432	5999	4044	4044	1616258	771	206.85	23.95	4.70	964
17	3353	3351	4432	1079	1079	1584022	793	32.19	24.53	5.93	992
18	4316	3208	3351	-965	965	1540503	805	22.36	24.39	4.65	1006
19	4479	4049	3208	-1271	1271	1545186	834	28.38	24.64	2.96	1042
20	5035	4610	4049	-986	986	1511442	843	19.58	24.34	1.76	1053
21	5604	5039	4610	-994	994	1482364	851	17.74	23.97	0.57	1064
22	5937	5525	5039	-898	898	1446756	854	15.12	23.51	-0.48	1067
23	5074	5538	5525	451	451	1384603	833	8.90	22.78	0.05	1042
24	5353	5455	5538	185	185	1320305	803	3.46	21.86	0.28	1003

- Para cada uno de los métodos seleccionados, se prepararon las fórmulas para la evaluación de los errores de pronóstico. Según sea el método, se agregaron otras columnas necesarias para los cálculos, tales como el nivel y la tendencia de la serie de tiempo.

Procedimiento utilizado

- Las medidas de error de pronósticos utilizadas fueron:
 - Error del pronóstico
 - Error absoluto del pronóstico.
 - Promedio de los cuadrados de errores (MSE).
 - Desviación absoluta media (MAD).
 - Error porcentual del pronóstico.
 - Error porcentual absoluto medio (MAPE).
 - Señal de rastreo (TS).
 - Desviación estándar.

Procedimiento utilizado

Autoguardado MP Proyecto final... Guardado en Este PC

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Automatizar Programador Ayuda PDFelement Power Pivot

Comentarios Compartir

Calibri 11 Fuente Alineación Número Estilos Celdas Edición

H6 =SUMA(\$F\$6:F6)/(A6-\$A\$5)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Denuncias																
2	Periodo	Denuncias	Lt	Ft	Et	At	MSE	MAD	%Error	MAPE	TSt	St					
3	1	5482															
4	2	5482															
5	3	5459	5474														
6	4	5815	5585	5474	-341	341	116054	341	5.86	5.86	-1.00	426					
7	5	6089	5788	5585	-504	504	184867	422	8.27	7.07	-2.00	528					
8	6	6044	5983	5788	-256	256	145147	367	4.24	6.12	-3.00	459					
9	7	6538	6224	5983	-555	555	185959	414	8.49	6.72	-4.00	518					
10	8	6219	6267	6224	5	5	148772	332	0.08	5.39	-4.97	415					
11	9	5935	6231	6267	332	332	142347	332	5.59	5.42	-3.97	415					
12	10	6401	6185	6231	-170	170	126156	309	2.66	5.03	-4.82	386					
13	11	5652	5996	6185	533	533	145898	337	9.43	5.58	-2.84	421					
14	12	5636	5896	5996	360	360	144087	340	6.39	5.67	-1.76	424					
15	13	6655	5981	5896	-759	759	187236	381	11.40	6.24	-3.55	477					
16	14	6597	6296	5981	-616	616	204710	403	9.34	6.52	-4.89	503					
17	15	4745	5999	6296	1551	1551	388118	498	32.69	8.70	-0.84	623					
18	16	1955	4432	5999	4044	4044	1616258	771	206.85	23.95	4.70	964					
19	17	3353	3351	4432	1079	1079	1584022	793	32.19	24.53	5.93	992					
20	18	4316	3208	3351	-965	965	1540503	805	22.36	24.39	4.65	1006					
21	19	4479	4049	3208	-1271	1271	1545186	834	28.38	24.64	2.96	1042					
22	20	5035	4610	4049	-986	986	1511442	843	19.58	24.34	1.76	1053					
23	21	5604	5039	4610	-994	994	1482364	851	17.74	23.97	0.57	1064					
24	22	5937	5525	5039	-898	898	1446756	854	15.12	23.51	-0.48	1067					
25	23	5074	5538	5525	451	451	1384603	833	8.90	22.78	0.05	1042					
26	24	5353	5455	5538	185	185	1320305	803	3.46	21.86	0.28	1003					
27	25	5130	5100	5455	315	315	1266021	780	6.14	21.11	0.60	876					

Tabla dinámica Resumen métodos Descriptiva Random Tree Multilayer Perceptron SMOreg Random Forest Promedio móvil Suavizam

Accesibilidad: es necesario investigar

- Los métodos clásicos seleccionados de pronóstico fueron:
 - Promedios móviles
 - Suavizamiento exponencial
 - Holt
 - Winter (HoltWinters)
 - Regresión lineal.
- En el caso de los promedios móviles, se eligió una ventana de 3 meses para el cálculo de cada promedio.

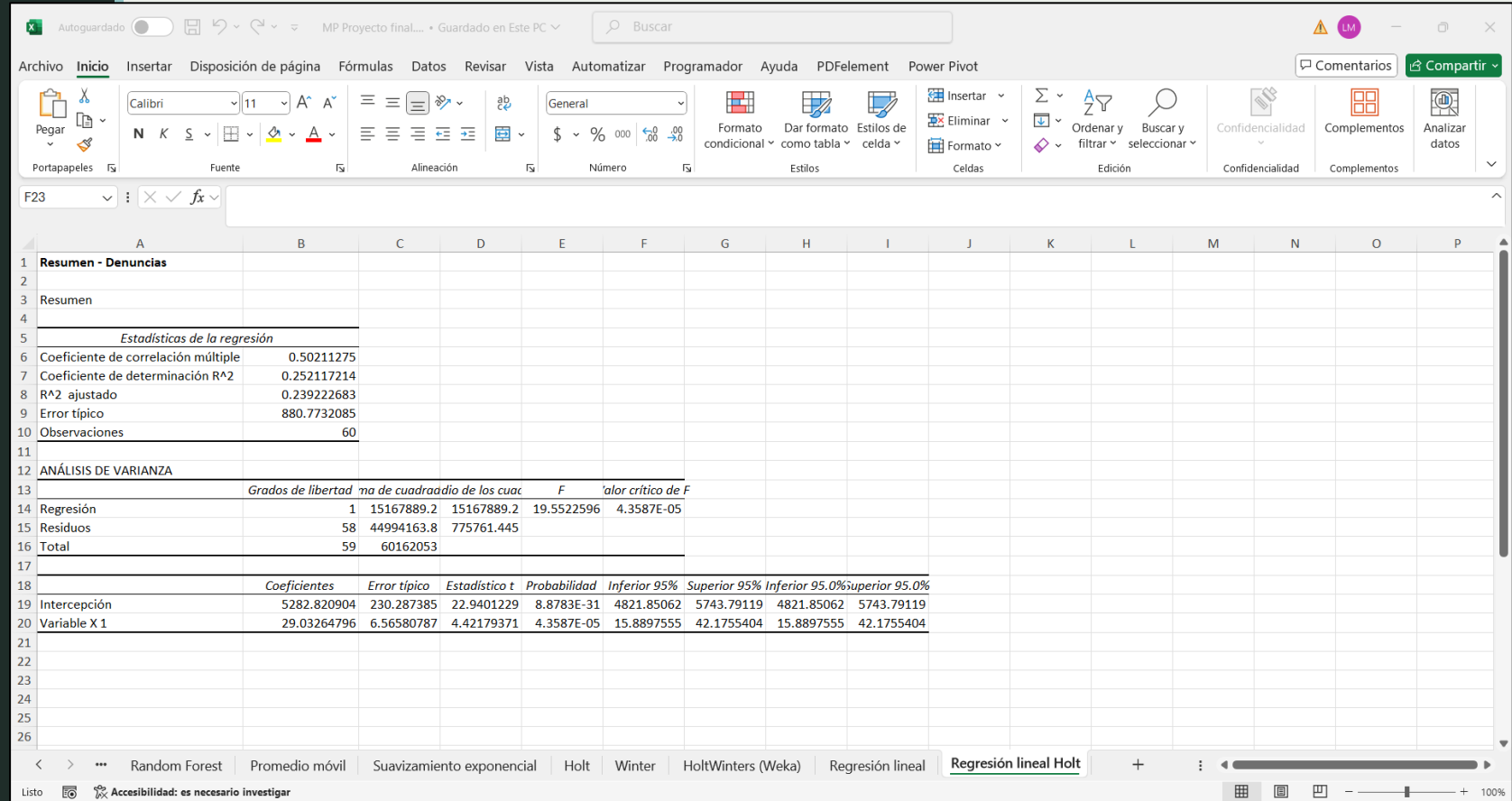
Procedimiento utilizado

Excel interface showing a spreadsheet with data for exponential smoothing. The spreadsheet includes columns for Periodo, Denuncias, Lt, Ft, Et, At, MSE, MAD, %Error, MAPE, TSt, and St. The 'Suavizamiento exponencial' tab is selected at the bottom.

Periodo	Denuncias	Lt	Ft	Et	At	MSE	MAD	%Error	MAPE	TSt	St
0		6168									
1	5482	5767	6168	686	686	471031	686	12.52	12.52	1.00	858
2	5482	5600	5767	285	285	275992	485	5.19	8.85	2.00	607
3	5459	5517	5600	141	141	190617	371	2.58	6.76	3.00	463
4	5815	5692	5517	-298	298	165099	352	5.12	6.35	2.31	440
5	6089	5924	5692	-397	397	163659	361	6.53	6.39	1.15	452
6	6044	5994	5924	-120	120	138771	321	1.98	5.65	0.93	401
7	6538	6313	5994	-544	544	161167	353	8.32	6.03	-0.70	441
8	6219	6258	6313	94	94	142117	320	1.51	5.47	-0.48	401
9	5935	6069	6258	323	323	137905	321	5.44	5.46	0.53	401
10	6401	6263	6069	-332	332	135148	322	5.19	5.44	-0.50	402
11	5652	5905	6263	611	611	156833	348	10.82	5.93	1.29	435
12	5636	5748	5905	269	269	149812	342	4.78	5.83	2.10	427
13	6655	6279	5748	-907	907	201612	385	13.63	6.43	-0.49	481
14	6597	6465	6279	-318	318	194441	380	4.82	6.32	-1.33	475
15	4745	5458	6465	1720	1720	378730	470	36.25	8.31	2.58	587
16	1955	3407	5458	3503	3503	1122041	659	179.19	18.99	7.15	824
17	3353	3375	3407	54	54	1056212	624	1.62	17.97	7.65	780
18	4316	3926	3375	-941	941	1046675	641	21.79	18.18	5.97	802
19	4479	4250	3926	-553	553	1007676	637	12.34	17.87	5.15	796
20	5035	4709	4250	-785	785	988120	644	15.60	17.76	3.87	805
21	5604	5233	4709	-895	895	979170	656	15.96	17.67	2.44	820
22	5937	5645	5233	-704	704	957180	658	11.86	17.41	1.36	823
23	5074	5311	5645	571	571	929750	654	11.26	17.14	2.24	818
24	5323	5323	5311	12	12	881895	638	0.70	16.46	2.26	786

- En el caso del suavizamiento exponencial, se utilizó el complemento Solver de Microsoft Excel con el fin de obtener el valor de la constante de suavizamiento α .

Procedimiento utilizado



The screenshot shows the Microsoft Excel interface with a regression analysis table. The table is located in the range A13:P20. The table has 8 columns: 'Grados de libertad', 'suma de cuadrados', 'F', 'valor crítico de F', 'Coeficientes', 'Error típico', 'Estadístico t', and 'Probabilidad'. The rows include 'Regresión', 'Residuos', 'Total', 'Intercepción', and 'Variable X 1'. The 'Regresión' row shows values for 'Grados de libertad' (1), 'suma de cuadrados' (15167889.2), 'F' (19.5522596), and 'valor crítico de F' (4.3587E-05). The 'Residuos' row shows values for 'Grados de libertad' (58), 'suma de cuadrados' (44994163.8), and 'F' (775761.445). The 'Total' row shows values for 'Grados de libertad' (59) and 'suma de cuadrados' (60162053). The 'Intercepción' row shows values for 'Coeficientes' (5282.820904), 'Error típico' (230.287385), 'Estadístico t' (22.9401229), 'Probabilidad' (8.8783E-31), 'Inferior 95%' (4821.85062), 'Superior 95%' (5743.79119), 'Inferior 95.0%' (4821.85062), and 'Superior 95.0%' (5743.79119). The 'Variable X 1' row shows values for 'Coeficientes' (29.03264796), 'Error típico' (6.56580787), 'Estadístico t' (4.42179371), 'Probabilidad' (4.3587E-05), 'Inferior 95%' (15.8897555), 'Superior 95%' (42.1755404), 'Inferior 95.0%' (15.8897555), and 'Superior 95.0%' (42.1755404).

	Grados de libertad	suma de cuadrados	F	valor crítico de F	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95.0%	Superior 95.0%
Regresión	1	15167889.2	15167889.2	19.5522596	4.3587E-05							
Residuos	58	44994163.8	775761.445									
Total	59	60162053										
Intercepción					5282.820904	230.287385	22.9401229	8.8783E-31	4821.85062	5743.79119	4821.85062	5743.79119
Variable X 1					29.03264796	6.56580787	4.42179371	4.3587E-05	15.8897555	42.1755404	15.8897555	42.1755404

- Para el método Holt, en primer lugar, se utilizó la opción de regresión del complemento de análisis de datos de Microsoft Excel.

Procedimiento utilizado

Autoguardado

MP Proyecto final... Guardado en Este PC

Buscar

Archivo

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

PDFelement

Power Pivot

Comentarios

Compartir

Pegar

Portapapeles

Calibri

11

A

A

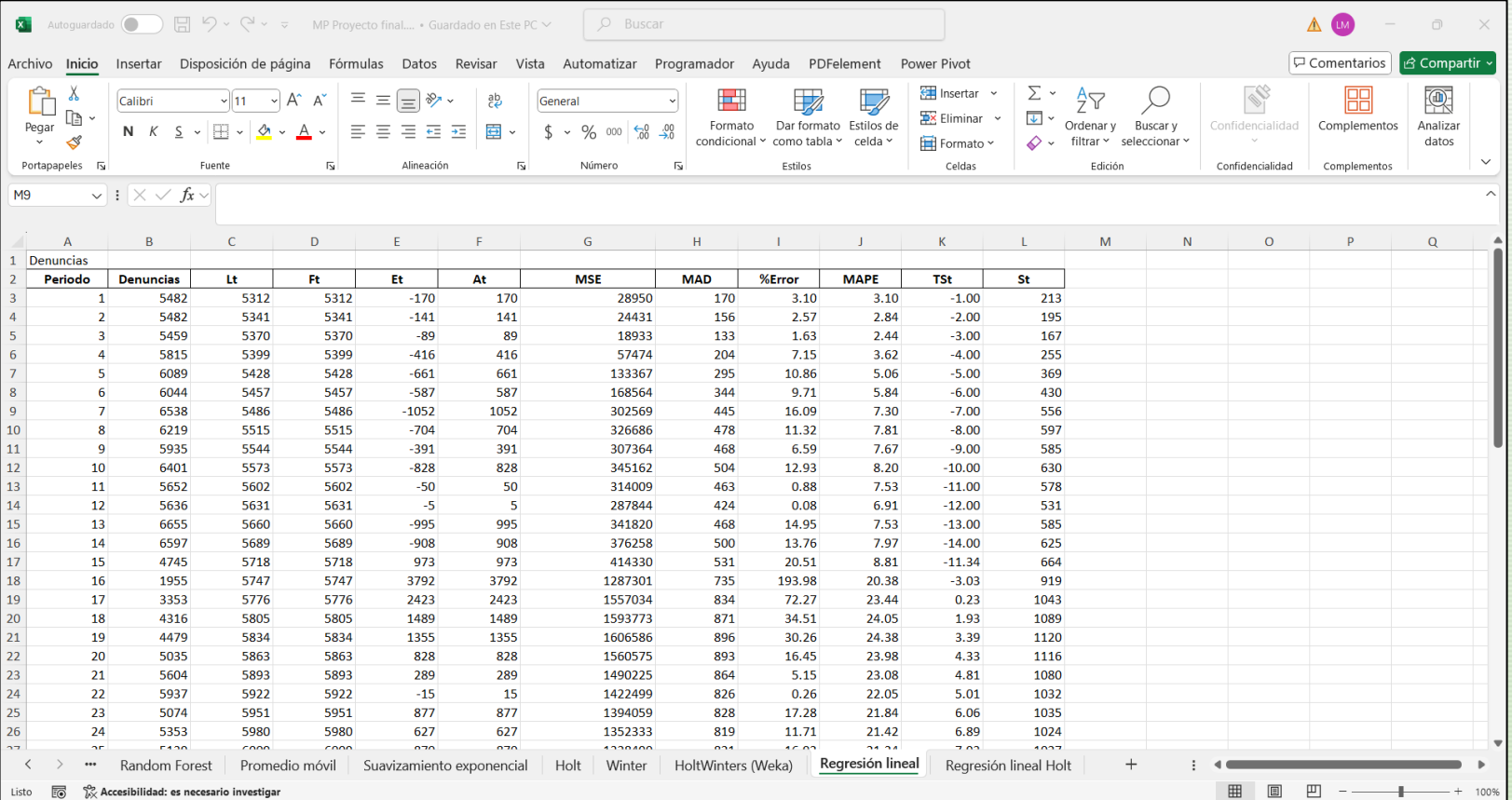
N

K

U

- Con el informe de resultados de la regresión se completó la hoja de cálculo para este método. Igualmente se utilizó el complemento Solver para encontrar el valor óptimo para la constante de suavizamiento α .

Procedimiento utilizado



Periodo	Denuncias	Lt	Ft	Et	At	MSE	MAD	%Error	MAPE	TSt	St
1	5482	5312	5312	-170	170	28950	170	3.10	3.10	-1.00	213
2	5482	5341	5341	-141	141	24431	156	2.57	2.84	-2.00	195
3	5459	5370	5370	-89	89	18933	133	1.63	2.44	-3.00	167
4	5815	5399	5399	-416	416	57474	204	7.15	3.62	-4.00	255
5	6089	5428	5428	-661	661	133367	295	10.86	5.06	-5.00	369
6	6044	5457	5457	-587	587	168564	344	9.71	5.84	-6.00	430
7	6538	5486	5486	-1052	1052	302569	445	16.09	7.30	-7.00	556
8	6219	5515	5515	-704	704	326686	478	11.32	7.81	-8.00	597
9	5935	5544	5544	-391	391	307364	468	6.59	7.67	-9.00	585
10	6401	5573	5573	-828	828	345162	504	12.93	8.20	-10.00	630
11	5652	5602	5602	-50	50	314009	463	0.88	7.53	-11.00	578
12	5636	5631	5631	-5	5	287844	424	0.08	6.91	-12.00	531
13	6655	5660	5660	-995	995	341820	468	14.95	7.53	-13.00	585
14	6597	5689	5689	-908	908	376258	500	13.76	7.97	-14.00	625
15	4745	5718	5718	973	973	414330	531	20.51	8.81	-11.34	664
16	1955	5747	5747	3792	3792	1287301	735	193.98	20.38	-3.03	919
17	3353	5776	5776	2423	2423	1557034	834	72.27	23.44	0.23	1043
18	4316	5805	5805	1489	1489	1593773	871	34.51	24.05	1.93	1089
19	4479	5834	5834	1355	1355	1606586	896	30.26	24.38	3.39	1120
20	5035	5863	5863	828	828	1560575	893	16.45	23.98	4.33	1116
21	5604	5893	5893	289	289	1490225	864	5.15	23.08	4.81	1080
22	5937	5922	5922	-15	15	1422499	826	0.26	22.05	5.01	1032
23	5074	5951	5951	877	877	1394059	828	17.28	21.84	6.06	1035
24	5353	5980	5980	627	627	1352333	819	11.71	21.42	6.89	1024

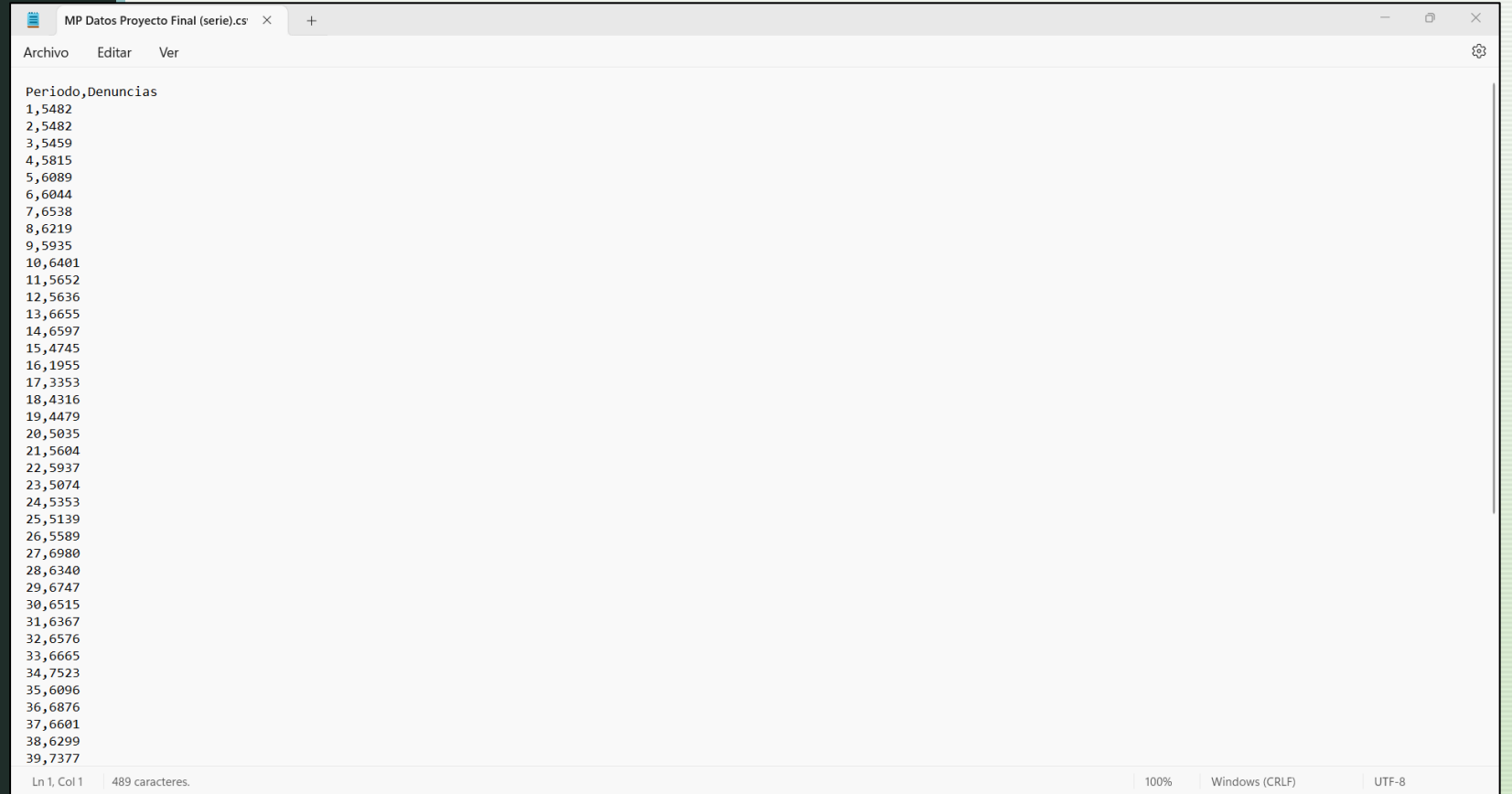
- Igualmente, el informe de resultados de la regresión sirvió para completar la hoja de cálculo para el método de regresión lineal.

Procedimiento utilizado



- Para ejecutar los algoritmos de machine learning, se descargó e instaló el software Weka en su versión básica (llamada *stable*).
- Con el fin de obtener los pronósticos, se activó la opción “Forecast” en Weka. Esta opción no está activa por omisión. Es necesario descargar un complemento para Weka e instalarlo.
- Para poder realizar pronósticos en Weka es necesario cargar un archivo con la serie de tiempo que será evaluada. El software solamente acepta ciertos formatos de archivo, entre ellos el formato CSV y un formato propio con extensión .arff (Attribute Relation File Format).

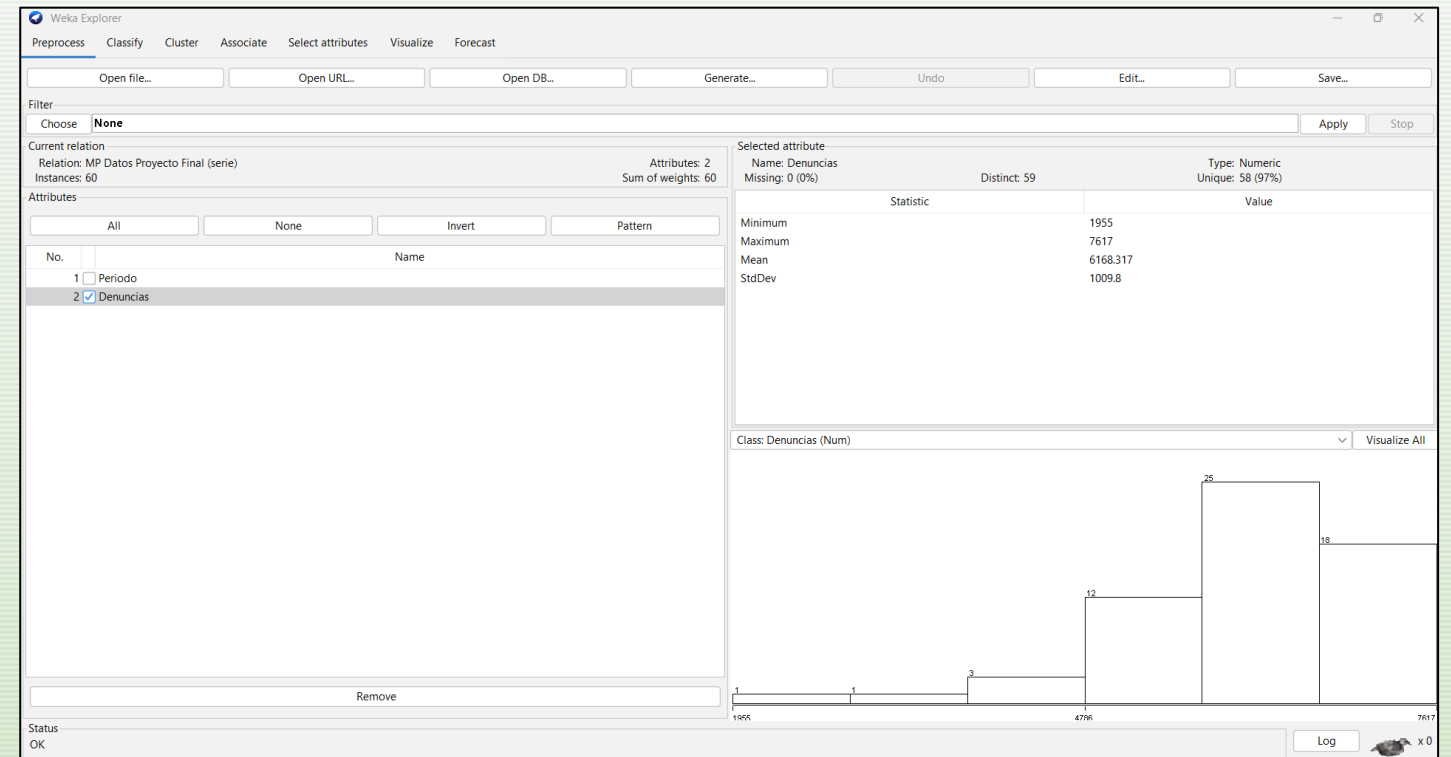
Procedimiento utilizado



```
MP Datos Proyecto Final (serie).cs × +
Archivo Editar Ver
Periodo,Denuncias
1,5482
2,5482
3,5459
4,5815
5,6089
6,6044
7,6538
8,6219
9,5935
10,6401
11,5652
12,5636
13,6655
14,6597
15,4745
16,1955
17,3353
18,4316
19,4479
20,5035
21,5604
22,5937
23,5074
24,5353
25,5139
26,5589
27,6980
28,6340
29,6747
30,6515
31,6367
32,6576
33,6665
34,7523
35,6096
36,6876
37,6601
38,6299
39,7377
Ln 1, Col 1 489 caracteres. 100% Windows (CRLF) UTF-8
```

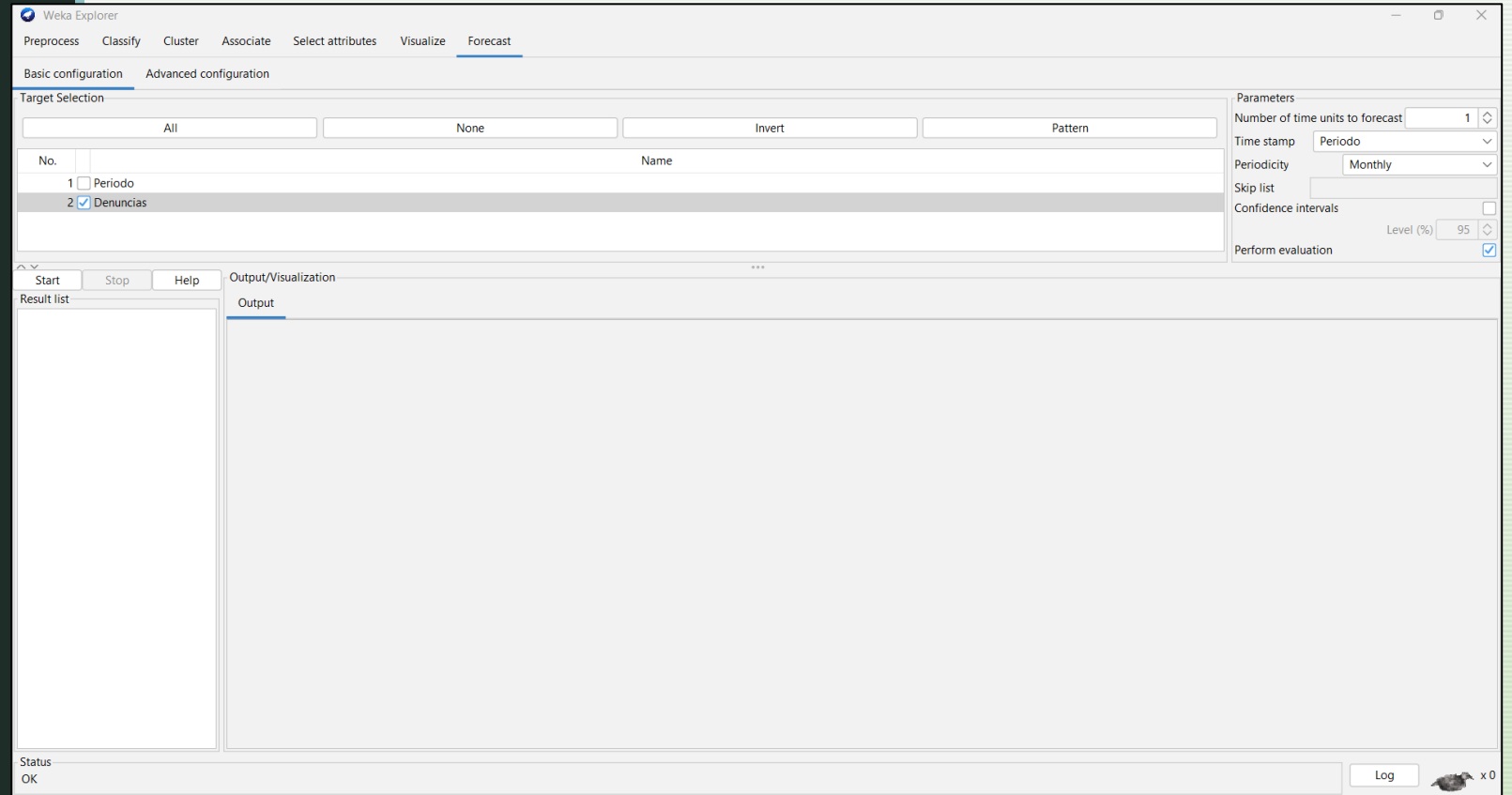
- Por lo anterior, se copió a un archivo nuevo de Excel el resultado de la tabla dinámica agrupada por mes.
- Este archivo de Excel se grabó como formato CSV con el fin de importarlo en el software Weka, que tiene la capacidad de convertir archivos CSV al formato .arff.

Procedimiento utilizado



- Se ejecutó la aplicación Weka. Apareció un menú inicial, en el cual se seleccionó la opción “Explorer”. Con esto, se abrió una nueva ventana, que es el explorador de datos del software.
- Se seleccionó la opción “Open file...” para cargar el archivo CSV creado anteriormente con los datos de la serie de tiempo. El software Weka automáticamente analiza el contenido del archivo y elabora una estadística descriptiva simple, con un histograma, el mínimo, el máximo, la media y la desviación estándar de los datos.

Procedimiento utilizado

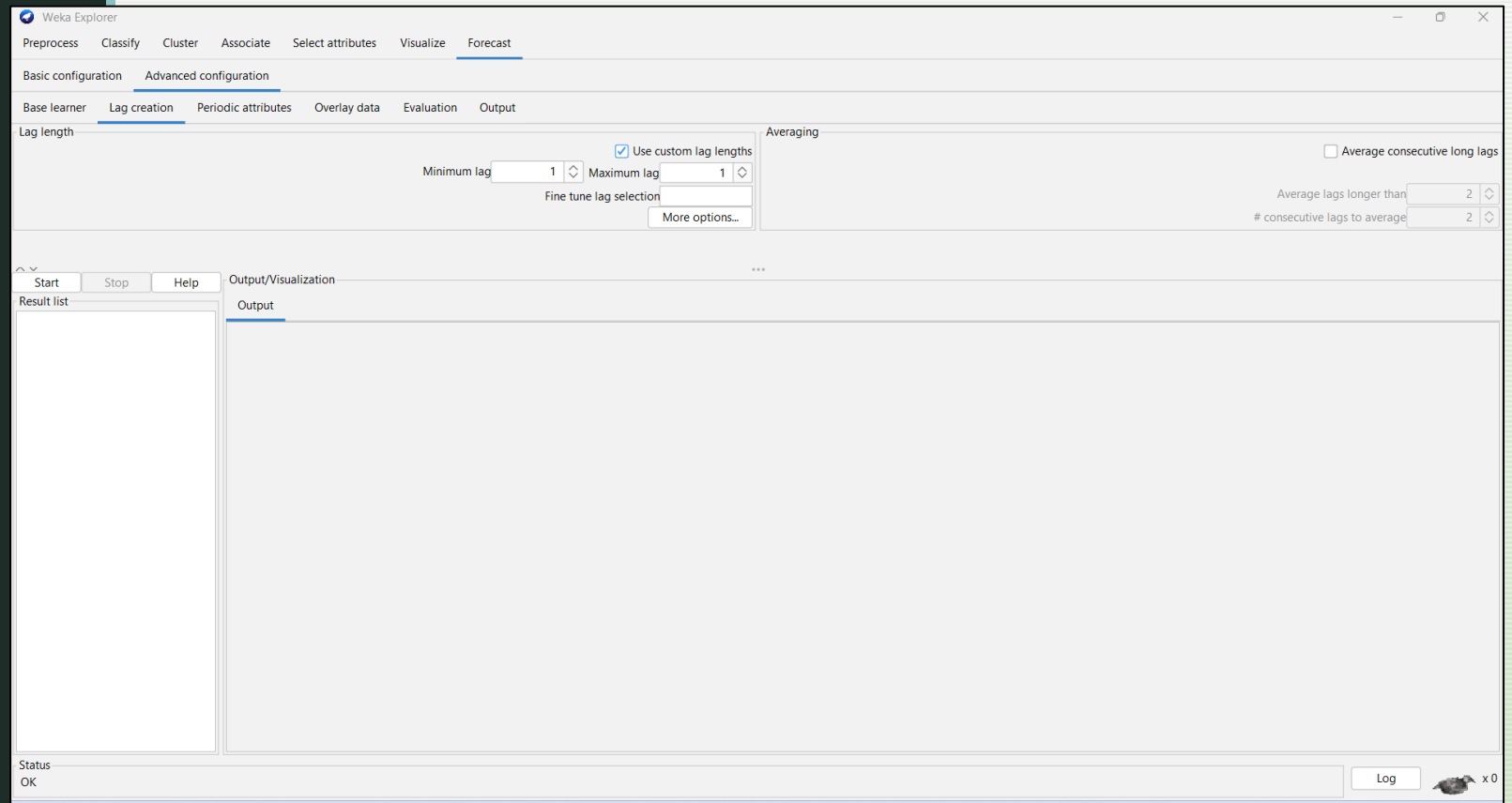


- En el menú del explorador se seleccionó la opción “Forecast” para hacer los pronósticos.
- El software Weka contiene varios parámetros y configuraciones que pueden ser modificados según la necesidad de cada análisis que se vaya a realizar.

Procedimiento utilizado

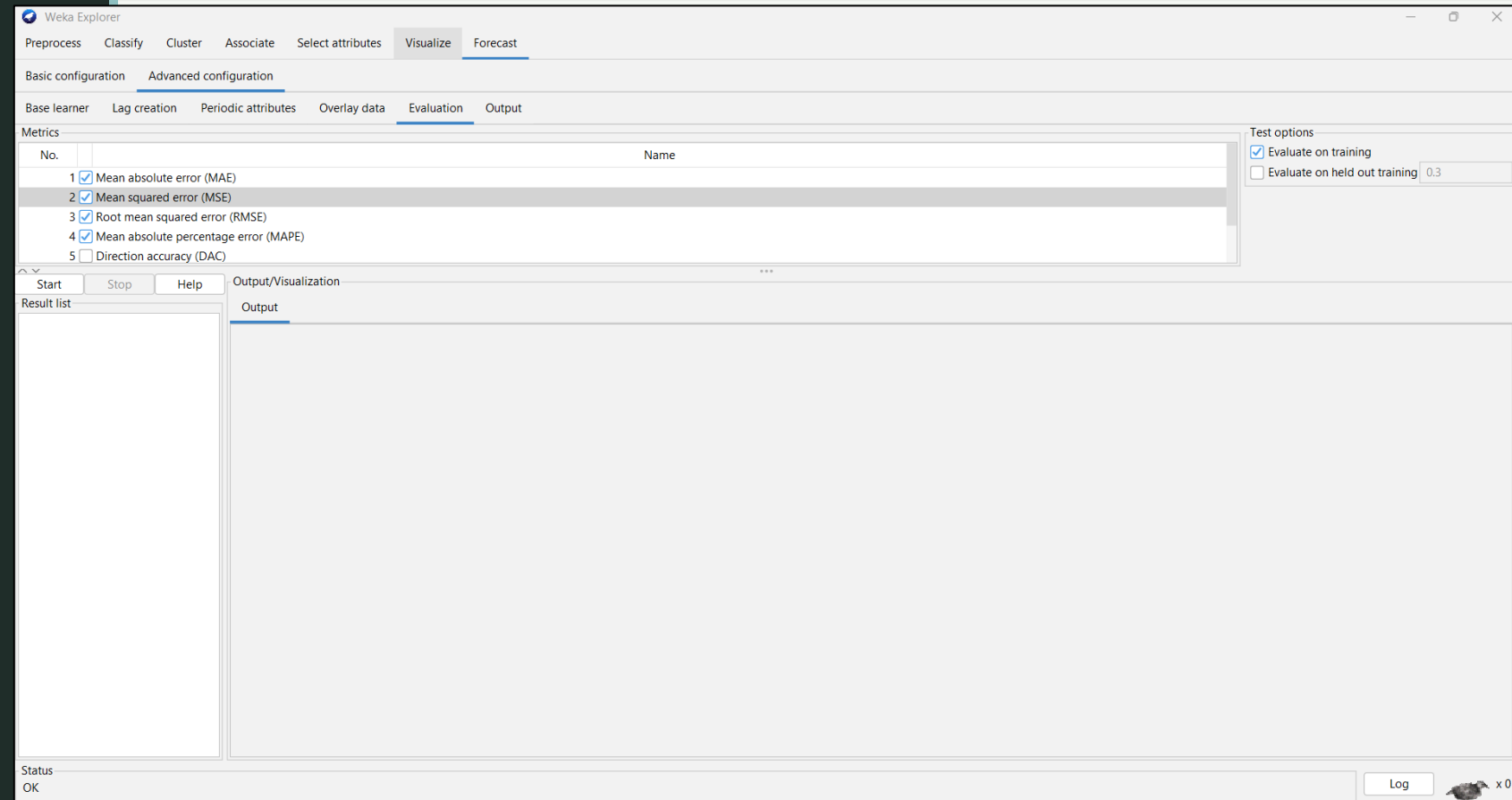
- Para la serie de tiempo de las denuncias presentadas se ajustaron varios parámetros.
 - Time-stamp: Se eligió el campo Período, para que fuera la guía del software para distinguir las unidades de tiempo.
 - Periodicity: Se seleccionó la opción “Monthly” porque los datos fueron agrupados por mes.
 - Perform evaluation: Se marcó la casilla para obtener un informe de evaluación de los pronósticos.
- En la sección “Target selection” se elige el campo Denuncias, que es el campo de donde se obtendrán los pronósticos.

Procedimiento utilizado



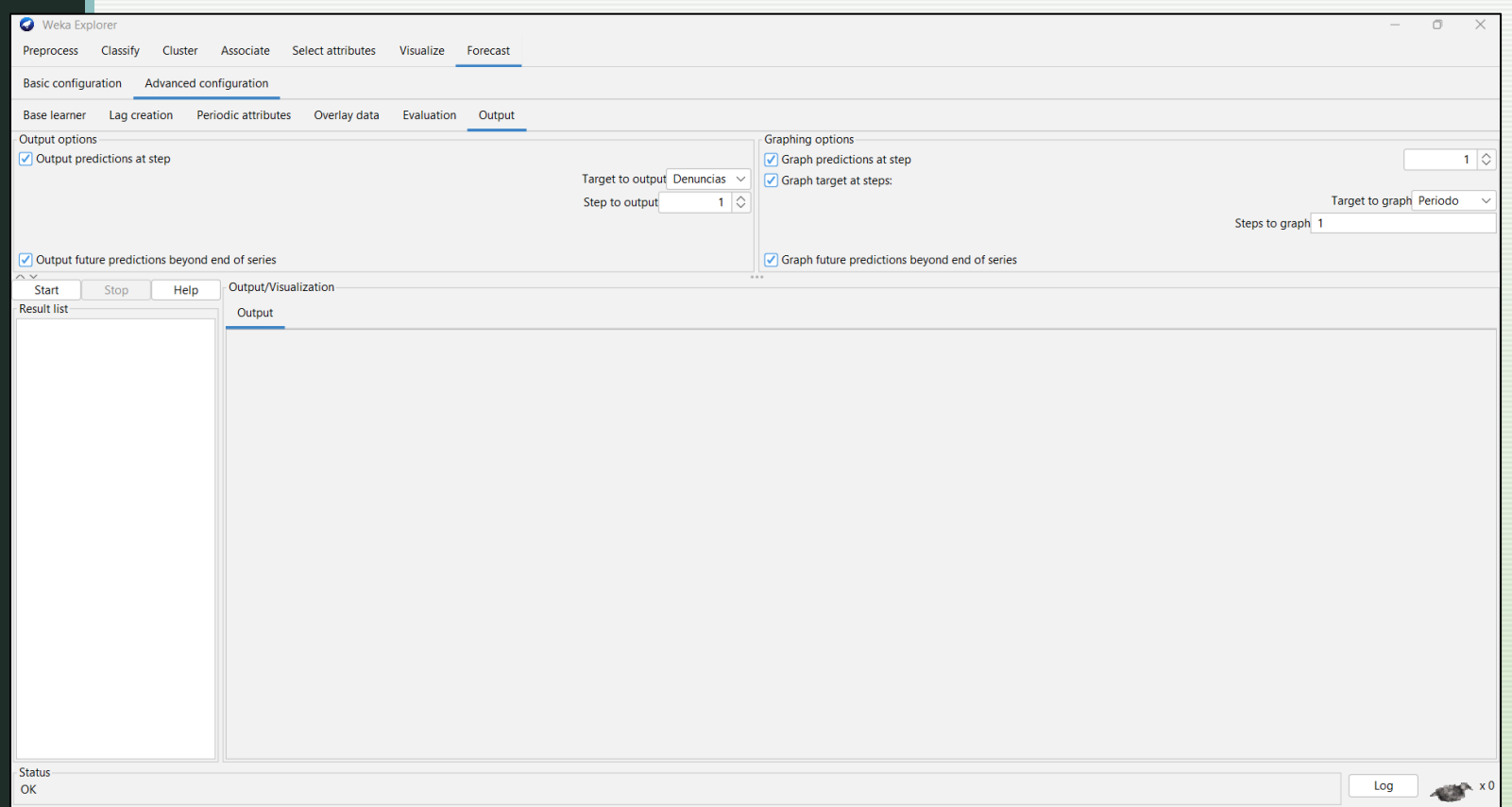
- En la pestaña “Advanced configuration” se modificaron otros valores para definir la modalidad de cálculo y la salida de los resultados.
- En la pestaña “Lag creation”
 - Se marcó la casilla “Use custom lag length”.
 - Se modificó a 1 el valor de “Maximum lag”.

Procedimiento utilizado



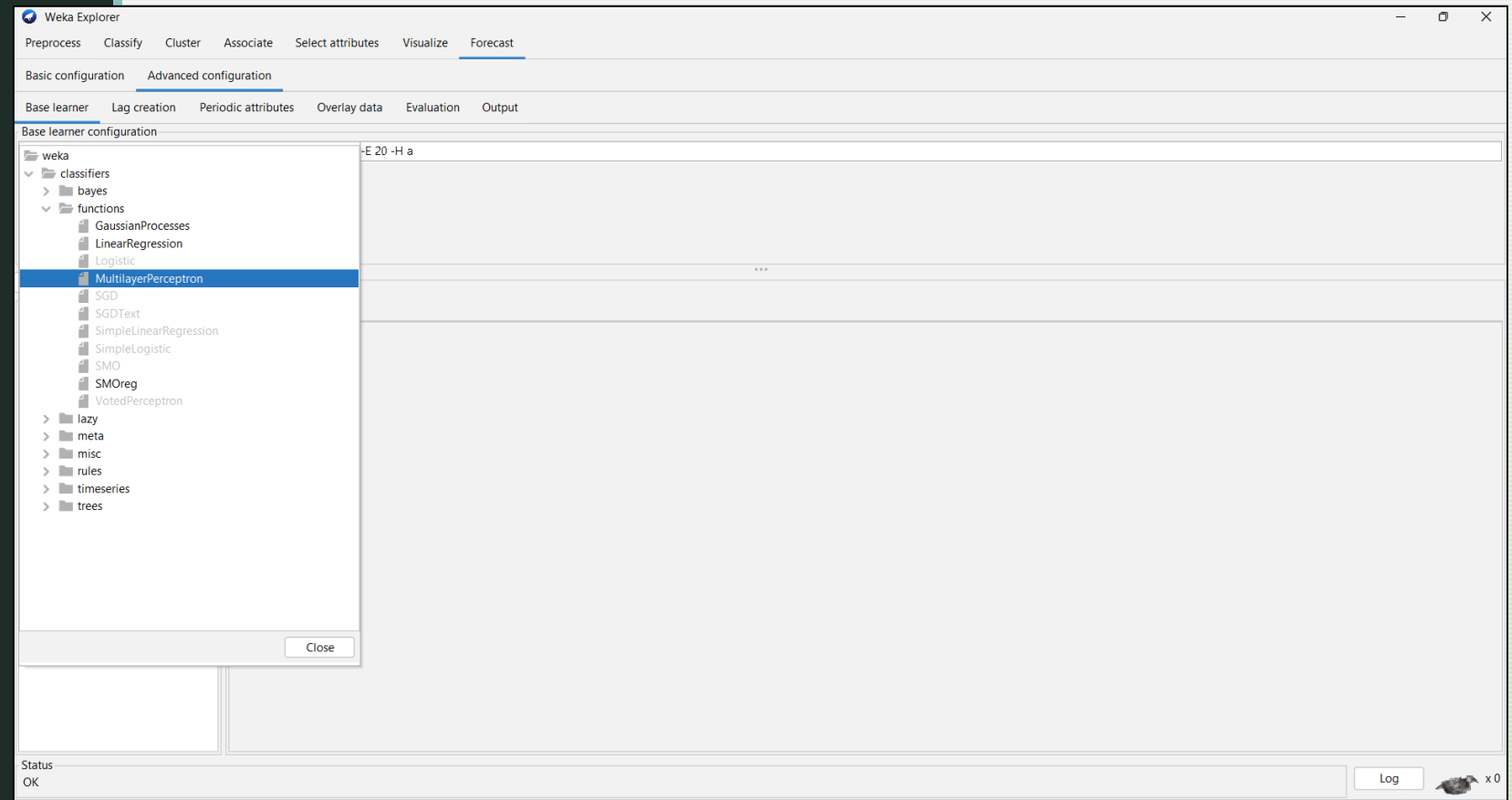
- En la pestaña “Evaluation” aparecieron por omisión dos casillas marcadas que son las de “Mean Absolute Error (MAE)” y “Root mean squared error (RMSE)”. Además, se marcaron las casillas de “Mean squared error (MSE)” y “Mean absolute percentage error (MAPE)”. Esto tiene la finalidad de obtener medidas de error que sean comparables con las de los métodos clásicos.
- Se dejó marcada la casilla de “Evaluate on training”.

Procedimiento utilizado



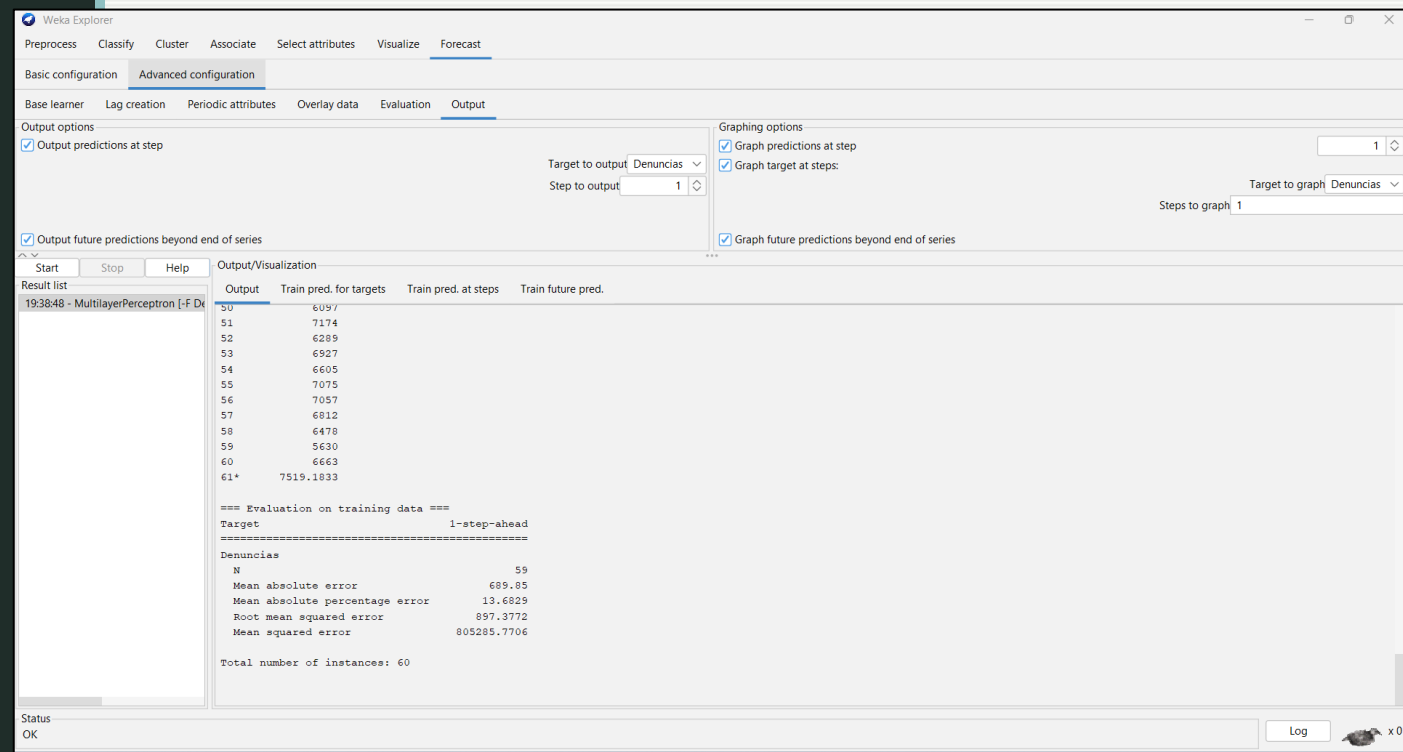
- En la pestaña “Output” se realizaron los siguientes cambios:
 - Se marcó la casilla “Output predictions at step”.
 - En el cuadro “Target to output” se seleccionó el campo Denuncias.
 - Se marcó la casilla “Graph predictions at step”.
 - Se marcó la casilla “Graph target at steps” y en el cuadro “Target to graph” se seleccionó el campo Denuncias.

Procedimiento utilizado



- En la pestaña “Base learners” se hace clic en el botón “Choose”, la cual permite elegir los algoritmos de machine learning que se ejecutan con los datos del archivo.

Procedimiento utilizado



- Con el botón “Start” se ejecutaron los siguientes algoritmos:
 - HoltWinters
 - SMOreg
 - Multilayer Perceptron
 - Random Forest
 - Random Tree
- Al presionar el botón “Start” se produce un informe de resultados en la ventana “Output/Visualization”. Para cada algoritmo se generó un informe y los gráficos solicitados en la configuración del software.

Procedimiento utilizado

- Los informes de Weka con las medidas de error permiten comparar los algoritmos seleccionados. Por otra parte, los gráficos permiten observar la precisión de los pronósticos por cada algoritmo.
- Para cada algoritmo se grabaron los resultados de los pronósticos y de las medidas de error en archivos de texto.
- Se transcribieron los resultados de los algoritmos al archivo consolidado del proyecto en Microsoft Excel. Cada algoritmo tiene su hoja de cálculo.
- En el archivo consolidado de Excel se preparó un cuadro resumen con los resultados de las medidas de error de pronósticos tanto de los métodos clásicos como de los algoritmos de machine learning.

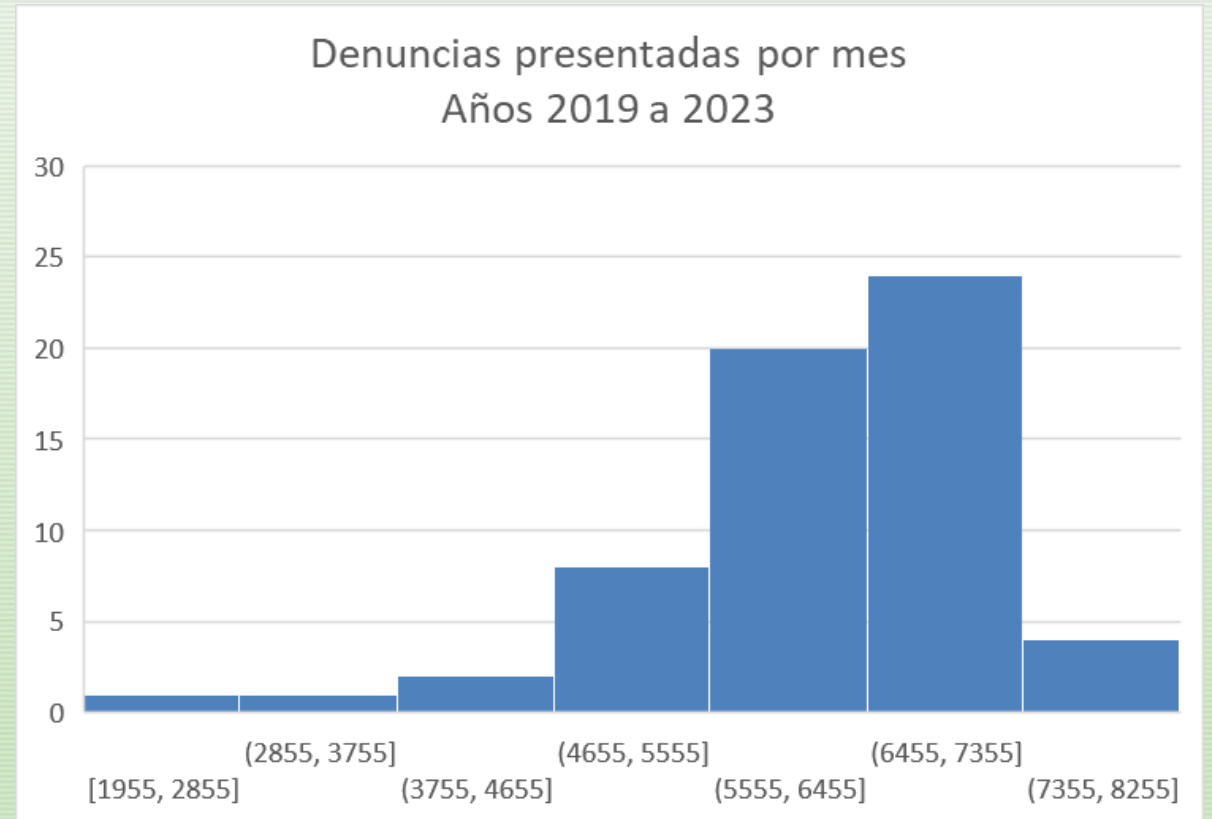
Resultados



ANÁLISIS EXPLORATORIO DE DATOS

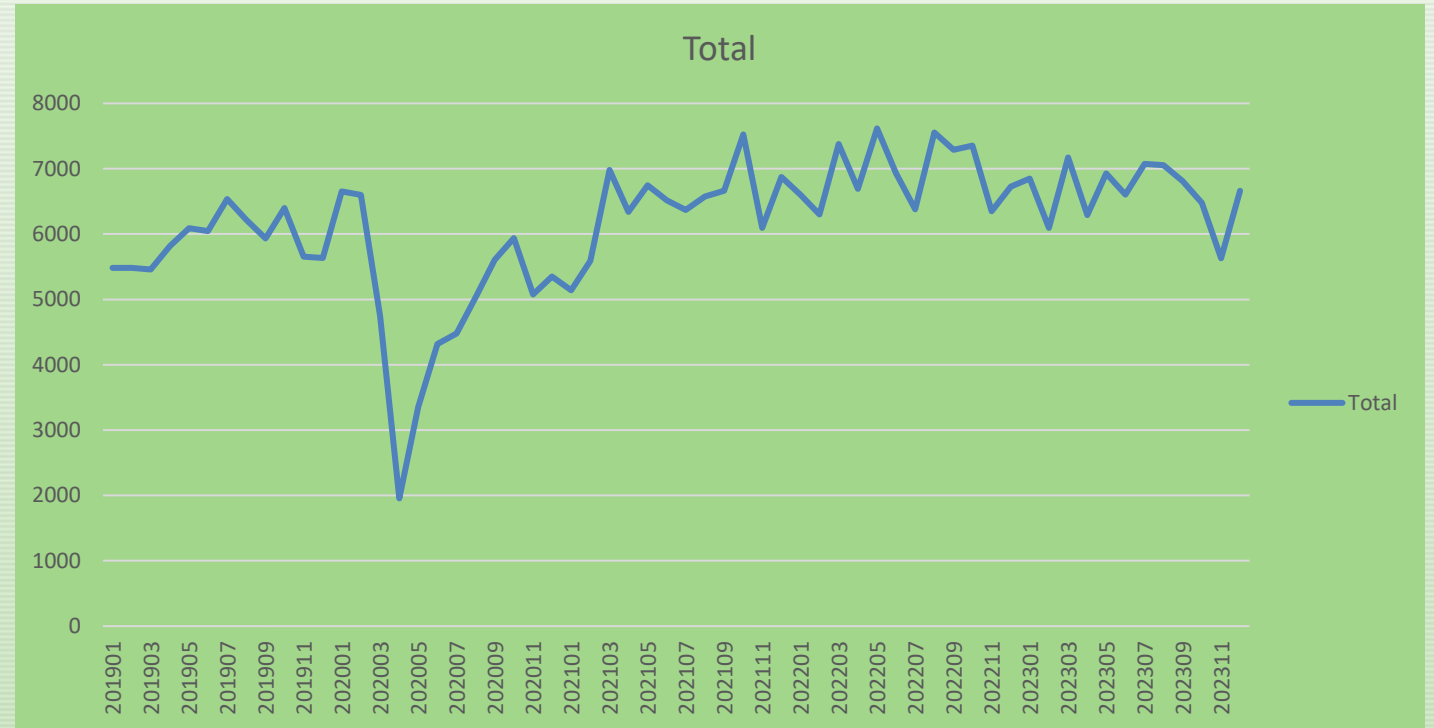
Resultados

<i>Denuncias</i>	
Media	6168.32
Error típico	130.36
Mediana	6374
Moda	5482
Desviación estándar	1009.80
Varianza de la muestra	1019695.81
Curtosis	4.59
Coeficiente de asimetría	-1.67
Rango	5662
Mínimo	1955
Máximo	7617
Suma	370099
Cuenta	60



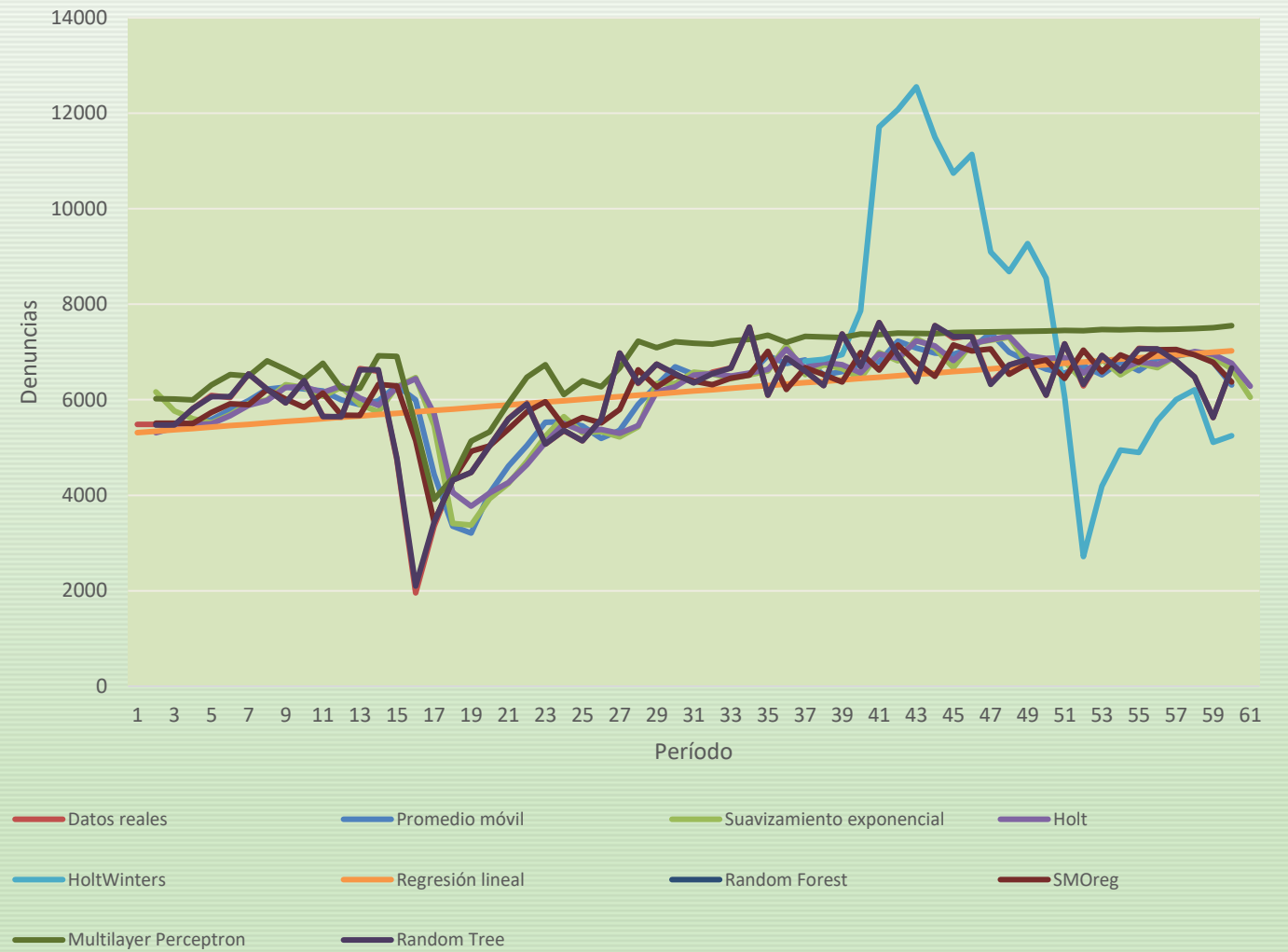
Resultados

SERIE DE TIEMPO PARA LAS DENUNCIAS PRESENTADAS



Resultados

Comparación de modelos predictivos para la presentación
de denuncias criminales en Panamá entre 2019 y 2023
Primer y tercer distritos judiciales



Resultados

COMPARACIÓN DE MÉTODOS DE PRONÓSTICO

Método de pronóstico	MAD	MAPE	Señal de rastreo inferior	Señal de rastreo superior	Desviación estándar
Promedio móvil	582	12.37	-6.07	5.93	728
Suavizamiento exponencial	526	10.71	-3.57	7.65	658
Holt	508	10.66	-7.46	7.27	635
Regresión lineal	615	12.87	-14.00	8.62	768
HoltWinters	2263	33.10	0.83	13.84	2829
SMOreg	463	9.29	-8.46	8.25	579
Multilayer Perceptron	690	13.68	10.00	54.56	862
Random Forest	171	3.61	-1.86	12.12	214
Random Tree	11	0.29	-2.52	21.59	14

Resultados

COMPARACIÓN DE MÉTODOS DE PRONÓSTICO

Método de pronóstico	MAD	MAPE	Señal de rastreo inferior	Señal de rastreo superior	Desviación estándar
Promedio móvil	582	12.37	-6.07	5.93	728
Suavizamiento exponencial	526	10.71	-3.57	7.65	658
Holt	508	10.66	-7.46	7.27	635
Regresión lineal	615	12.87	-14.00	8.62	768
HoltWinters	2263	33.10	0.83	13.84	2829
SMOreg	463	9.29	-8.46	8.25	579
Multilayer Perceptron	690	13.68	10.00	54.56	862
Random Forest	171	3.61	-1.86	12.12	214
Random Tree	11	0.29	-2.52	21.59	14

Conclusiones



Conclusiones

- Los tres algoritmos que tuvieron mejor resultado son:
 - Random Tree
 - Random Forest
 - SMOreg
- Una alternativa de pronóstico es el uso de cadenas de Markov, en las cuales se asignan probabilidades a cada uno de los valores posibles de pronóstico con la finalidad de obtener una mejor estimación.
- Otra alternativa de pronóstico para las denuncias presentadas es seleccionar el valor intermedio de los pronósticos de estos algoritmos.

Mes	Valor real	SMOreg	Random Forest	Random Tree	Pronóstico seleccionado
ene-2024	7316	6914	6380	5630	6380
feb-2024	6629	7058	6443	5630	6443
mar-2024	6608	7147	6310	5630	6310

Conclusiones

- Para el caso de la serie de tiempo de las denuncias presentadas ante el Ministerio Público de Panamá, ningún método o algoritmo pudo manejar con alto nivel de precisión los pronósticos. Una causa posible es la ausencia de patrones específicos en la serie de tiempo.
- La serie de tiempo tiene componentes de aleatoriedad que provocan variaciones que hasta ahora resultan difíciles de predecir más allá de los datos de entrenamiento del modelo. Posiblemente la aleatoriedad se deba al hecho de que la presentación de denuncias es influida por múltiples factores, tales como:
 - Crecimiento demográfico no planificado.
 - Falta de confianza en las instituciones.
 - Naturaleza de los delitos.
 - Influencia del crimen organizado.

Gracias

