



Aplicación de Machine Learning para el proceso de clasificación de datos asociados al centro médico Celina

Machine Learning application for the data classification process associated with the Celina medical center

Autores:

Espín Espín, Deysi Magaly¹

Vilcacundo Pérez, Susana Margoth²

Banda Casa, Mario Agustín³

RESUMEN

El presente proyecto de investigación tiene por objetivo la aplicación de la inteligencia artificial mediante el uso de Machine Learning para el proceso de clasificación de datos asociados al centro médico CMC, en el cual se aplica la metodología CRISP–DM que permite el análisis de datos para la obtención de información útil mediante un proceso de clasificación de las preferencias del servicio médico que se proporciona a los pacientes. A partir del análisis de la data generada por el centro médico durante los años 2021 y 2022. Se procede a la creación de modelos parametrizados para predicción para lo cual en su aplicación se encuentran las mencionadas preferencias entre, citas médicas con sus distintas especialidades, tratamientos de fisioterapia manual y tecnológica, y control de signos vitales; de donde se obtiene que la preferencia del cliente está en el servicio de citas médicas en el área de Ginecología y Obstetricia. Esta información resulta importante para identificar hacia donde se proyecta el giro de negocio del centro médico.

Palabras Clave: machine learning, aprendizaje automático, árbol de decisión, regresión logística, modelo CRISP–DM.

Recibido: 17/04/2023 **Aceptado:** 10/09/2023 **Publicado:** 24/09/2023

¹ Docente Pontificia Universidad Católica del Ecuador. Email: dmespin@puce.edu.ec ORCID: <https://orcid.org/0009-0001-6035-5087>

² Docente Universidad Técnica de Cotopaxi. Email: susana.vilcacundo6966@utc.edu.ec ORCID: <https://orcid.org/0000-0003-0502-2260>

³ Docente Universidad Técnica de Cotopaxi. Email: mario.banda@utc.edu.ec ORCID: <https://orcid.org/0000-0003-3950-1286>

ABSTARCT

The objective of this research project is the application of artificial intelligence through the use of Machine Learning for the data classification process associated with the CMC medical center, in wich CRISP–DM methodology is applied that allows data analysis to obtain useful information through a preference classification process for the medical service provided to patients during the years 2021 and 2022. Parameterized models for prediction are created, for wich, in the application you will find the aforementioned preferences between, medical appointments with their different specialties, manual and technological physiotherapy treatments, and vital signs monitoring; from where, we obtain that the client's preference is in the medical appointment service, in the area of Gynecology and Obstetrics. This information is important to identify where the business of the medical center is projected.

Keywords: machine learning, decision tree, logistic regression, CRISP–DM model.

INTRODUCCIÓN

En la actualidad, la información de las organizaciones, instituciones y empresas, se convierte en uno de los activos más importantes dentro de su actividad organizacional, además, debido al aumento de productos relacionados con Internet y al incremento del poder de procesamiento de las computadoras, la información se va almacenando de forma masiva en distintos repositorios, ya sea en la “nube”, o de manera local en un servidor de información. Esta facilidad de disponer del uso de este valioso recurso como es la “información”, y de aplicar estrategias tecnológicas sobre esta, crea la posibilidad de establecer una ventaja competitiva frente a las demás organizaciones.

Con los progresivos avances en el área de la informática y la computación, el aprendizaje automático, uno de los grandes avances de los dispositivos

computacionales, ha adquirido varias formas de empleo, incluido el procesamiento de señales de manera general y, el procesamiento de imágenes y videos de manera particular. Aunque el aprendizaje automático de las máquinas suele asociarse a métodos de gran potencia de cálculo, mediante una compleja estructuración informática, en la actualidad existe un gran interés por incorporar algoritmos a dispositivos y sensores conectados a Internet de las Cosas. (Estrella, 2020). Existen múltiples opciones que determinan las ventajas de emplear la Inteligencia Artificial, una de ellas es el uso de la Clasificación mediante la implementación de modelos predictivos.

“La clasificación en el aprendizaje automático se refiere a entrenar un modelo en un conjunto de datos etiquetados para categorizar puntos de datos existentes o

para clasificar nuevos puntos de datos. Se utilizan técnicas de modelado predictivo como la clasificación” (AEFOL, 2022).

La organización adecuada de la información es importante para las organizaciones e instituciones de carácter público o privado; tal es el caso de la información que almacena el centro médico CMC de la ciudad de Quito, en donde se presenta esta necesidad y que es en donde se emplea el modelo de Machine Learning para la clasificación de la información del

METODOLOGÍA

La metodología es el conjunto de herramientas, programas de aplicación, métodos, técnicas, entre otros; que permiten llevar a cabo un proceso determinado y alcanzar una gama de objetivos. Para el desarrollo de proyectos para Minería de Datos se dispone de varias opciones; en el presente proyecto, se hace uso de una metodología especializada denominada CRISP-DM, cuyo acrónimo significa “Cross Industry Standar Procces for Data Mining”; o Proceso estándar transversal de la industria para minería de datos (IBM, 2021).

Para la programación de cada uno de los modelos se utiliza el lenguaje de programación Python, debido a su versatilidad y facilidad de uso.

METODOLOGÍA CRISP-DM

“Es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema”. (Galán, 2015, pag. 21)

centro médico que permita predecir o clasificar y conocer el porcentaje de demanda existente en el centro médico tanto en citas médicas como en tratamiento de fisioterapia y signos vitales.

Con los resultados encontrados, entonces, se genera un escenario favorable en el CMC para que los directivos tomen las mejores decisiones y también para conocer a ciencia cierta, hacia qué aspectos del servicio se está direccionando el giro de negocio.

La metodología CRISP-DM forma parte de una secuencia de fases para un plan de minería de datos que son:

Comprensión del negocio: Es la que nos permite entender el giro de negocio para ello se necesita realizar las algunas actividades o tareas como son, Identificación del problema, determinación de los objetivos y evaluación del estado actual. (Rollings, 2015, p. 4)

Comprensión de los datos: Esta fase permite comprender como esta los datos para ser analizados para ello se tiene las siguientes actividades o tareas que son: Recolección de datos, descripción de los datos y exploración de los datos. (IBM, 2021)

Preparación de los datos. Esta fase es la que consume más tiempo, ya que requiere la selección de datos que se transformarán, en función de los resultados de las fases anteriores, se utilizarán adecuadamente en la fase de modelado. Para ello, se realiza las

siguientes actividades o tareas: limpieza de datos, crear indicadores y Transformación de los datos. (IBM, 2021)

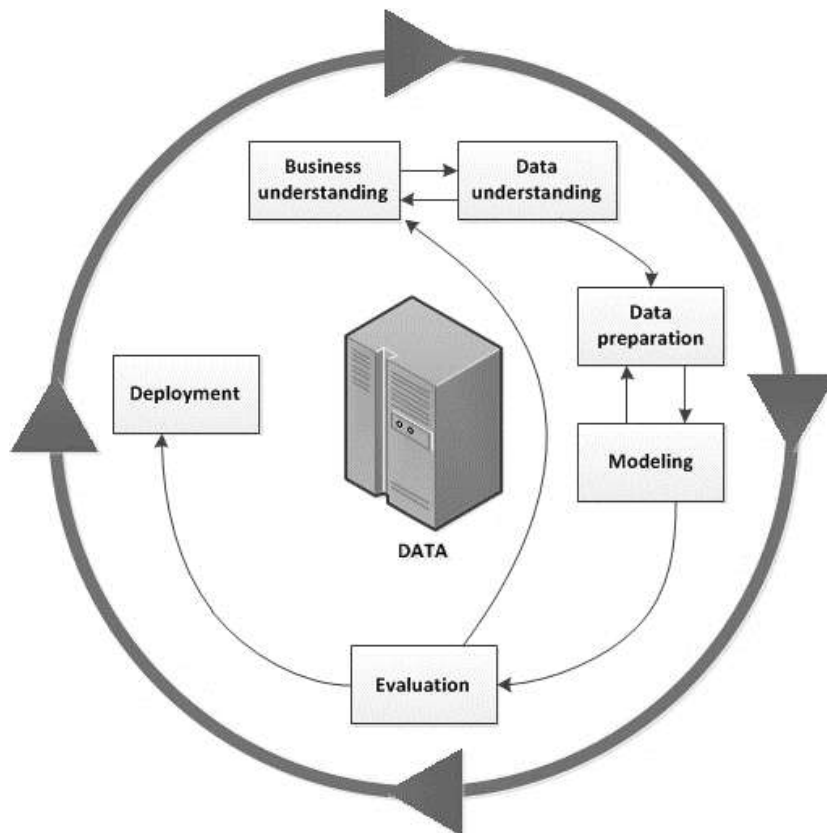
Modelado: Esta etapa es propiamente la minería de datos, para ello se sugiere: Seleccionar el método para el modelado, Selección de los datos de prueba y obtención del Modelo. (Rollings, 2015, p. 4)

Evaluación: En esta fase se evaluará la calidad del modelo en base al análisis de métricas y estadísticas y comparará los resultados con los resultados anteriores,

todo dependiendo de los resultados, y esta vez se determina o de acuerdo a la última parte del proceso. o Queda por volver a los pasos anteriores o volver a empezar desde el primer paso (Rollings, 2015)

Despliegue: Documentar los resultados obtenidos en las fases para que la persona encargada pueda leer con claridad y también todas las fases sean documentadas, con eso permitirá realizar una revisión final de todos los procesos que se obtiene (IBM, 2021).

Figura 1
Fases del modelo de referencia CRISP-DM



Nota. Tomado de IBM, (2021)

RESULTADOS

Entre los principales resultados se tiene.

Comprensión del negocio

El centro médico CMC brinda los servicios en diferentes especialidades a cada uno de sus clientes. Es aquí en donde se ve la necesidad de predecir cuál es el área con mayor demanda. Se tiene en laboratorio o consulta médica, en fisioterapia, fisioterapia manual, fisioterapia tecnológica y en signos vitales.

Determinar objetivos de minería de datos

Obtener información valiosa que permita realizar un análisis comparativo dentro de los data_set de citas médicas, tratamiento de fisioterapia y signos vitales para verificar cuál de los grupos es más frecuente.

Analizar la base obtenida de los años 2021 a 2022 con 24.546 registros.

Establecer si los tipos de problemas de minería de datos que se pueden analizar son predicciones tanto en citas médicas, tratamiento de fisioterapia y signos vitales

Tabla 1

Plan de proyecto. Estimación de tiempo de ejecución de las fases de la metodología CRISP-DM

Fase	Tiempo	Recursos
Compresión del negocio	1 semanas	Analista
Compresión de los datos	3 semanas	Analista
Preparación de datos	5 semanas	Ejecución de minería de datos, análisis base de datos.
Modelado	2 semanas	Ejecución de minería de datos, análisis base de datos.
Evaluación	1 semanas	Analista
Despliegue	1 semanas	Consultoría de minería de datos análisis base de datos.

Nota. Elaboración propia

Tabla 2

Comprensión de los datos

id	Fecha_c	Hora	Detalle	Tipo cita	Historia	id_doc	Estado	Abono_c	Saldo_c	Total_c	Fecha_ci	
0	1.0	2022-07-21	14:00	Cita médica: Consultas Traumatología	Traumatología	15070.0	19.0	Atendido - Cancelado	0.0	0.0	25.0	Nat
1	2.0	2022-07-21	17:00	Cita médica: Consultas Medicina Interna	Medicina Interna	3398.0	16.0	Atendido - Cancelado	15.0	10.0	25.0	Nat
2	3.0	2022-07-21	17:30	Cita médica: Consultas MEDICINA INTERNA	Medicina Interna	15061.0	16.0	Atendido - Cancelado	0.0	0.0	25.0	Nat
3	4.0	2022-07-22	15:00	Cita médica: Consultas Traumatología	Traumatología	15080.0	19.0	Atendido - Cancelado	0.0	0.0	25.0	7/21/2020 16:18:26 PM

4	5.0	2022-07-22	17:00	Cita médica: Consultas Medicina Interna	Medicina Interna	15084.0	16.0	Atendido - Cancelado	0.0	0.0	25.0	7/22/2020 14:10:42 PM
---	-----	------------	-------	--	---------------------	---------	------	----------------------------	-----	-----	------	-----------------------------

Nota. Elaboración propia

Tabla 3

Datos de signos vitales agrupados

	ID_APA R2	HISTOR IA	TEMP	TEM	PESO	FRECU_CAR DIA	TENSION_ARTE RIAL	PERIME_FALI CO	FREC_RE SP	TALL A	FECHA2	HORA
0	26475	13824	NaN	37. 0	72. 2	73	162/95	0.0	0.0	165. 0	13/01/20 20	15:22: 00
1	20011	12535	NaN	38. 0	70. 8	122	111/84	0.0	25.0	164. 0	14/05/20 19	8:00:0 0
2	27049	14177	NaN	37. 0	71. 9	72	140/86	0.0	18.0	164. 0	19/03/20 20	12:49: 00
3	509	5659	NaN	37. 0	56. 1	90	109/79	0.0	0.0	161. 0	16/01/20 16	15:07: 00
4	37931	15565	NaN	36. 3	59. 2	55	102/55	77.0	18.0	160. 0	30/11/20 20	10:22: 00

Nota. Elaboración propia

Tabla 3

La siguiente imagen presenta datos de exámenes fisioterapia (Data Set Tratamiento _ Fisioterapia)

	id_exa	HISTORIA	id_es	id_grado	Diag_doc	Diag_tera	Exa_comple	Deporte	Mati_cons	Exa_fisico
0	238	11194	10	2	Fractura de femur superior coxoartrosis	Coxoartrosis osteoartrosis edema de miembros inferiores	RX	Ninguno	Dolor intenso en el área lumbar y glútea	Pct presenta dolor intenso en el área lumbar y glúteo
1	229	11194	6	4	Fractura de femur superior coxoartrosis	Coxoartrosis osteoartrosis edema de miembros inferiores	RX	Ninguno	Dolor intenso en el área lumbar y glútea	Pct presenta dolor intenso en el área lumbar y glúteo
2	239	11194	8	4	Fractura de femur superior coxoartrosis	Coxoartrosis osteoartrosis edema de miembros inferiores	RX	Ninguno	Dolor intenso en el área lumbar y glútea	Pct presenta dolor intenso en el área lumbar y glúteo
3	308	11294	10	3	Esguince residivante	Esguince residivante y fascitis plantar del pie	RX	Ninguno	Dolor en el pie izquierdo	Palpación ul a nivel de tobillo izquierdo

4	310	11294	7	3	Esguince residivante	Esguince residivante y fascitis plantar del pie	RX	Ninguno	Dolor en el pie izquierdo	Palpación ul a nivel de tobillo izquierdo
---	-----	-------	---	---	----------------------	---	----	---------	---------------------------	---

Nota. Elaboración propia

Limpieza de datos

- Para limpiar en citas médica, se realizan las siguientes tareas:
- Colocar el rango de fecha del 2021 al 2022.
- Verificar los datos duplicados.
- Eliminación de datos duplicados.
- Verifica valores únicos.
- Cambiar al mismo nombre por minúsculas a mayúsculas y por faltas de ortografía.
- Agrupar la columna a predecir Tipo_cita a 0 las especialidades que pertenezca al área de laboratorio y 1 que pertenece a área consulta médica.
- Transformar las variables object a numérico.
- Diagrama de los mínimos y máximos para tener una idea cómo se comporta los datos.
- Verificar valores nulos.
- Diagrama de bigotes para ver si presenta outliers.

Modelado

En el data set de citas médicas describe 12 características, pero las que se va a usar son:

- Detalle: se detalla el contenido de la cita
- Tipo_cita: el tipo de especialista.
- HISTORIA = Número de historia del paciente.

- Estado = Es si está pendiente o no de atención.

- Año: año que es gestionado la cita.

En el data set tratamiento de fisioterapia describe 16 características, pero las que se va a usar son:

- Diag_doc: es el diagnóstico del doctor frente al paciente.
- Diag_tera = Diagnóstico del tratamiento.
- Trata_fisio1: tratamiento a realizar
- Num_Secciones: Cuantas secciones debe usar para la rehabilitación.
- Año: que gestiona el tratamiento

En el data set signos vitales describe 14 características, pero las que se va a usar son:

- TEMP = Temperatura tomada al paciente en 0 y 1.
- FRECU_CARDIA: Frecuencia cardiaca tomado al paciente.
- FRECU_RESP = Frecuencia respiratoria tomado al paciente.
- SATURA = saturación tomada al paciente.
- Año: Año que el paciente es atendido

Construcción del modelo

Configuración de parámetros: Después de la limpieza de los datos se escogen las variables y se establecen los parámetros de las citas médicas para su modelo; de igual manera se seleccionan las variables y se establecen los parámetros para

fisioterapia, y por último se selecciona las variables y se genera el modelo para los signos vitales.

Tabla 4

Distribución para la aplicación de los dos modelos en citas médica (Train y Test)

```
# split data
from sklearn.model_selection import train_test_split
x= df[[' HISTORIA', 'Detalle', 'Estado', 'AÑO']]
Y=df.Tipo_Cita
print(x)
print(y)
```

	HISTORIA	Detalle	Estado	AÑO
206	3002	231	0	2021
222	20	1219	0	2021
234	2892	1203	0	2021
235	10223	170	0	2021
236	10243	170	0	2021
4233	8617	1083	0	2021
4236	18451	1216	0	2021
4662	6272	856	0	2021
4666	18754	238	0	2021
4676	17157	47	0	2021

[1741 rows x 4 columns]

206	1
222	1
234	0
235	0
236	0
4233	1
4236	1
4662	1
4666	1
4676	1

Name: Tipo_Cita, Lenght: 1741, dtype: object

Nota. Elaboración propia, las variables agrupan sobre los ejes X e Y

Tabla 5

Distribución para la aplicación de los dos modelos en fisioterapia (Train y Test)

```
# Split data
from sklearn.model_selection import train_test_split
x= df[[' Diag_doc', 'Diag_tera', 'Num_secciones', 'AÑO']]
Y=df['Trata_fisio1']
print(x)
```



```
print(y)
```

	Diag_doc	Diag_tera	Num_secciones	AÑO
0	109	137	10	2022
1	109	137	10	2022
2	109	137	10	2022
3	98	212	10	2022
4	98	212	10	2022
1049	217	214	10	2022
1050	217	214	10	2022
1051	217	214	10	2022
1052	217	214	10	2022
1053	217	214	10	2022

[451 rows x 4 Columns]

0	1
1	1
2	1
3	0
4	0
1049	1
1050	1
1051	1
1052	1
1053	1

Name: Trata_fisio1, Lenght: 451, dtype: object

Nota. Elaboración propia las variables agrupan sobre los ejes X e Y.

Tabla 6

Distribución para la aplicación de los dos modelos en signos vitales (Train y Test)

```
# Split data
```

```
from sklearn.model_selection import train_test_split
```

```
x= df[['FRECU_CARDIA', 'FREC_RESP', 'SATURA', 'AÑO']]
```

```
Y=df['TEMP']
```

```
print(x)
```

```
print(y)
```

	FRECU_CARDIA	FREC_RESP	SATURA	AÑO
23	104	20	91	2021
25	88	95	95	2021
27	92	78	91	2021
36	91	0	97	2021
43	63	19	91	2022
16034	100	20	89	2022
16035	105	20	90	2022
16036	78	20	94	2022

18612	101	0	87	2022
18625	82	0	90	2021
[3231 rows x 4 columns]				
23	1			
25	0			
27	0			
36	1			
43	1			
16034	0			
16035	0			
16036	0			
18612	0			
18625	0			
Name: TEMP, LENGHT: 3221, dtype: float64				

Nota. Elaboración propia, las variables agrupan sobre los ejes X e Y.

Regresión Logística

Se aplica el modelo de regresión logística múltiple y el modelo decisión tree donde se pueden ver los valores que arrojan en la

aplicación del modelo y las gráficas del modelo de decisión tree.

Resultado regresión logística múltiple.

En la siguiente tabla se puede observar la aplicación de la librería para el procesamiento del modelado de donde se

obtiene los valores de predicción en los datos de prueba en logística múltiple.

Tabla 7

Modelado de datos.

```

from sklearn import linear_model
Classifier = linear_model.LogisticRegression()
Classifier.fit(X_train, y_train)
#make predictions for the test
# make predictions for the test data
y_pred = Classifier.predict(X_test)
print (list(y_test))
print (list(y_pred))
[1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

```



```

'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '1', '1', '1']
    
```

Nota. Elaboración propia

Resultado de decisión tree en signos vitales

En la siguiente tabla se observa el uso de la librería para el procesamiento del modelado se imprime los valores de

predicción para datos de prueba en signos vitales.

Tabla 9

Uso de librería

```

from sklearn.tree import DesicionTreeClassifier
#Classifier = DesicionTreeClassifier()
Classifier = DesicionTreeClassifier(max_depth=8)
Classifier.fit(X_train, y_train)
# make predictions for test data
y_pred1 = Classifier.predict(X_test)
print(y_pred1)
[0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
    
```

Nota. Elaboración propia

Evaluación de resultados

La evaluación del modelo de clasificación del data set citas médicas que está dividido en consultas médicas y laboratorio, se obtuvo una precisión del 91.98% con el modelo de regresión logística múltiple y

93.70% con el modelo de decisión tree, de esta forma el segundo modelo obtuvo mejor resultado de predicción. La evaluación del modelo de clasificación para tratamiento de fisioterapia en tratamiento

manual y tecnológico, se obtuvo una precisión del 74.73% con el modelo de regresión logística múltiple y 83.52% con el modelo de decisión tree, de esta forma el segundo modelo obtuvo mejor resultado de predicción.

La evaluación del modelo de clasificación para signos vitales clasificado en

emergencia y no emergencia, se obtuvo una precisión del 89.95% con el modelo de regresión logística múltiple y 89.34% el modelo de decisión tree, de esta forma el primer modelo obtuvo mejor resultado de predicción.

Resumen entre regresión logística múltiple y decisión tree de evaluación de matriz de confusión.

El siguiente cuadro muestra una matriz de confusión en donde se puede valorar cuál

de los modelos de clasificación presenta mejor condición a ser tomado en cuenta.

Tabla 10
Matriz de confusión

Data set	Regresión logística múltiple	Decisión tree
Citas médicas	[[0 28] [0 321]] Accuracy: 91.98%	[[7 21] [1 320]] Accuracy: 93.70%
Tratamiento de fisioterapia	[[68 0] [23 0]] Accuracy: 74.73%	[[65 3] [12 11]] Accuracy: 83.52%
Signos vitales	[[582 0] [65 0]] Accuracy: 89.95%	[[575 7] [62 3]] Accuracy: 89.34%

Nota. Elaboración propia

Presentación Final

Este estudio muestra el diferente análisis realizado tanto en consultas médicas laboratorio, tratamiento manual, tratamiento tecnológico y finalmente signos vitales de emergencia y no

emergencia, se muestra que en citas médicas la tiene más acogida, entre los tres data set analizados, es la de consulta médica en GINECOLOGÍA- OBSTETRICIA.

En el segundo data set de tratamiento de fisioterapia se escoge el total de datos que tiene el data set porque en los años 2021-2022 representaba muy pocos datos por lo que se incorpora el año 2019, como resultado se refleja con más frecuente la

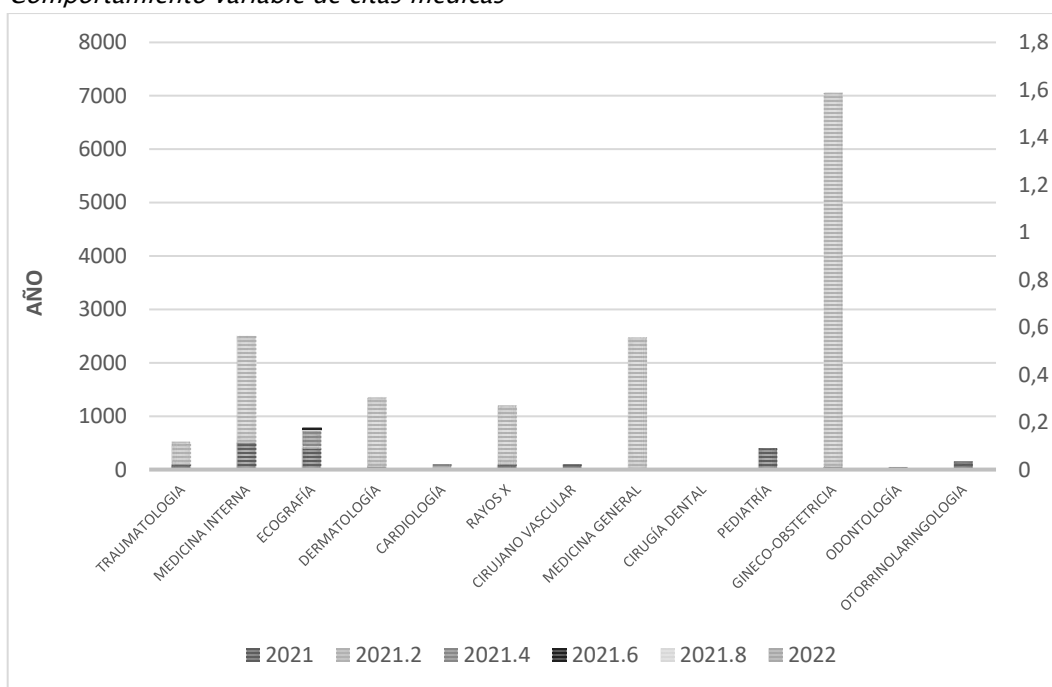
terapia tecnológica en el área de MAGNETOTERAPIA.

Y en el último data set, en la predicción, se obtuvo más información en la temperatura que es de emergencia en el rango del año 2021 al 2022.

Afluencia en las especialidades que oferta el centro médico

Figura 2

Comportamiento variable de citas médicas



Nota. Elaboración propia

Tratamiento de fisioterapia para cada paciente que es atendido.

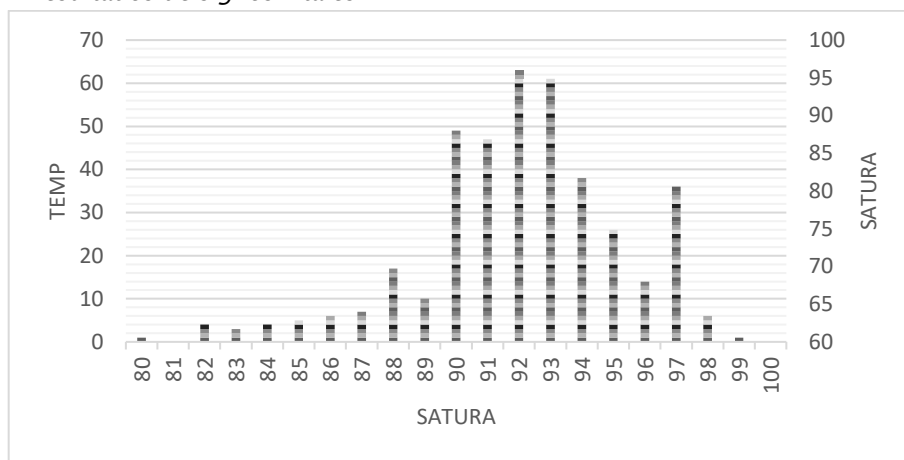
En el siguiente gráfico se observa el comportamiento variable de cada una de

las áreas de fisioterapia registradas durante los años 2019, 2021 y 2022.

En el siguiente gráfico se presenta los resultados de los signos vitales de pacientes que fueron atendidos en el año

2021 al 2022 con temperatura alta, baja y normal.

Figura 5
Resultados de signos vitales



Nota. Elaboración propia

DISCUSIÓN

El trabajo realizado de análisis de los datos del Centro médico CMC es un proceso que empieza mediante el reconocimiento de la actividad del giro de negocio que realiza, esto significa un valioso detalle para establecer los objetivos a alcanzar con el estudio. Esto concuerda con otros trabajos realizados en donde la actividad a la que se dedican se convierte en el eje fundamental para el tratamiento de la información.

Galán (2015), sigue el pie de la letra todo el proceso de implementación de la metodología CRISP-DM, y concuerda plenamente con que se debe conocer el contexto de la organización motivo de análisis. En el trabajo se establece como objetivos conocer mediante la predicción

los porcentajes de titulados, mediante una regresión logística, y por supuesto entender si se puede predecir un mejoramiento en el servicio de la institución.

Rojas (2016) Emplea Clusters para realizar la división y posterior agrupamiento mediante la conformación de grupos de pacientes asociados al cáncer, esto con la finalidad de incluir a las actividades realizadas por servidores para lograr un mejor servicio de salud.

Este último trabajo investigativo luego de aplicar el procedimiento necesario para conseguir mejores servicios, acude a los clusters para su clasificación, lo cual se considera un desperdicio de recursos ya

que el cluster está en un nivel mucho más profundo en el campo de la minería de datos, ya que el principal objeto de los clusters es la agrupación mediante una o algunas características proporcionadas en el modelo. Por otra parte, el uso de árbol de decisión resulta más fiable cuando se trata de clasificar bajo uno o varios criterios existentes, en donde el modelo será quien

decida de acuerdo con lo que encuentre en la data.

En el caso de la presente investigación, el modelo implementado adquiere la condición de clasificador de registros de actividad que permiten, establecer empleo de uno u otro servicio como preferencia de los usuarios pacientes.

CONCLUSIONES

A partir del trabajo de titulación del análisis de datos y la ejecución de la metodología CRISP-DM se puede concluir lo siguiente:

- El centro médico CMC maneja grandes cantidades de datos relacionados con pacientes, historiales médicos, diagnósticos, tratamientos y más. El proceso de clasificación y análisis de esta información puede mejorarse mediante el uso de inteligencia artificial, especialmente mediante el uso de métodos de aprendizaje automático con ayuda de la metodología CRISP-DM.
- En el caso del Centro Médico CMC, se pueden utilizar técnicas de machine learning para clasificar los datos de los pacientes en diferentes categorías según su diagnóstico, tratamiento y otros factores relevantes. Esto permitirá tener una mejor comprensión de los patrones y tendencias en los datos, a su vez

puede ayudar a mejorar la calidad y eficiencia de las atenciones médicas.

- La aplicación del análisis de ciencias de datos con la metodología CRISP_DM, en el centro médico ha demostrado ser un instrumento valioso en la toma de decisiones basadas en datos. La metodología CRISP-DM permite una evaluación objetiva y transparente de las soluciones propuestas por los algoritmos de inteligencia artificial de tal manera que se aumenta la confianza en los resultados obtenidos.
- El posicionamiento de la Inteligencia artificial, su uso y desarrollo en las áreas del conocimiento, actualmente se establece como un factor que marca una incidencia trascendental y progresiva de la realidad mundial, es decir, se acentúa un nuevo acontecimiento paradigmático de la tecnología computacional y los sistemas de información en donde,

la clasificación, la predicción, el reconocimiento de imágenes, entre otros, van siendo parte de la cotidianidad de la vida del ser humano, por lo que, las transformaciones dejan de a poco

de ser solo ciencia ficción para convertirse en una realidad moderna que trae intrínsecas muchas consecuencias a tener en cuenta en el futuro inmediato.

REFERENCIAS BIBLIOGRÁFICAS

AEFOL, E. (18 de 07 de 2022). 4 TIPOS DE TAREAS DE CLASIFICACIÓN EN EL MACHINE LEARNING. Obtenido de <https://elearningactual.com/4-tipos-de-tareas-de-clasificacion-en-el-machine-learning/>

Estrella, À. (2020). Aplicaciones basadas en aprendizaje automático (machine learnin) en plataforma de bajo consumo. Obtenido de https://oa.upm.es/66520/1/TFG_ALVARO_ESTRELLA_OLIVA.pdf

Galán, B. (2015). APLICACIÓN DE LA METODOLOGÍA CRISP-DM A UN PROYECTO DE MINERÍA DE DATOS EN EL ENTORNO UNIVERSITARIO. Trabajo de Grado. Universidad Carlos Tercero. Madrid. Obtenido de: https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf

IBM. (2021). Guía de CRISP-DM de IBM SPSS Modeler. Obtenido de https://inseaddataanalytics.github.io/INSEADAnalytics/CRISP_DM.pdf

Rojas, W. (2016). SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS

PARA APOYAR LA TOMA DE DECISIONES EN EL PROCESO DE CIRUGIA DEL HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS, EN LA CIUDAD DE MANIZALES. Obtenido de: https://repositorio.autonoma.edu.co/bitstream/11182/660/8/Soluci%C3%B3n_inteligencia_negocios_miner%C3%ADa_datos_toma_decisiones_proceso_cirug%C3%ADa_Hospital_Departamental_Universitario_Santa_Sof%C3%ADa_Caldas_ciudad_Manizales.pdf.

Rollings, B. (2015) Metodología Fundamental para la Ciencia de Datos. Obtenido de <https://www.ibm.com/downloads/cas/6RZMKDN8>