Machine Learning Models based in Supervised Learning for the Detection of Diabetes Mellitus in the City of Guayaquil.

Modelos de *Machine Learning* basados en Aprendizaje Supervisado para la Detección de Diabetes Mellitus en la Ciudad de Guayaquil.

Darwin Patiño-Pérez,Ph.D[1],Felicísimo Iñiguez-Muñoz,MSc[2],María Nivela-Cornejo,Ph.D[3],
Alicia Ruíz-Ramírez,MSc[3],Omar Otero-Agreda,MSc[3],Karla Game-Mendoza,MSc[4],Gladys Vinueza-Burgos,MSc[4],

[1]Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Física, Guayaquil,Ecuador,darwin.patinop@ug.edu.ec,
[2]Bitekso S.A, Guayaquil,Ecuador,bernardo.iniguez@bitekso.com,
[3]Universidad de Guayaquil, Facultad de Filosofía,Guayaquil,Ecuador, maria.nivelac@ug.edu.ec,
alicia.ruizr@ug.edu.ec,omar.oteroa@ug.edu.ec,

[4]Universidad Estatal de Milagro,Facultad de Ciencias de la Educación, Milagro,Ecuador,kgamen@unemi.edu.ec,
gvinuezab1@unemi.edu.ec

Abstract. – Due to the problems presented by public health centers in the city of Guayaquil in Ecuador, a group of specialists greeted the creation of a comprehensive diabetes diagnosis center, which, with an entrepreneurial initiative, is bringing together private health centers that are interested in providing a technologically assisted service. Using artificial intelligence tools, the comprehensive center detects whether a person has diabetes mellitus or not; Therefore, through this mechanism and applying the established triage, the person detected by the AI model is referred to the respective private health center that is part of the comprehensive center (entrepreneurship). Supervised learning was applied in the implementation of some conventional machine learning techniques, which were compared with an artificial neural network model to determine the model that can obtain the highest accuracy in learning diabetes diagnosis. As a result, it was obtained that the artificial neural network ANN reached 99.33% accuracy compared to the support vector machines SVM, which reached 98.6% accuracy in the classification of patients with diabetes mellitus, so that in view of the fact that the RNA model , is the one who has learned to detect diabetes with greater accuracy and is recommended to be used in the comprehensive diagnostic center, for rapid referral to a private health center that can apply rapid and effective treatment .

Keywords: Diabetes, Health Centers, Public Policies, Machine Learning, Supervised Learning.

Resumen. – Debido a los problemas que presentan los centros de salud públicos en la ciudad de Guayaquil en Ecuador, un grupo de especialistas en salud crearon un centro integral de diagnóstico de diabetes el mismo que con una iniciativa de emprendimiento está agrupando a centros de salud privados que estén interesados en brindar un servicio asistido tecnológicamente. Mediante el uso de herramientas de inteligencia artificial el centro integral detecta si una persona tiene diabetes mellitus o no; por lo que mediante este mecanismo y aplicando el triage establecido, se derivada a la persona detectada por el modelo de IA, al respectivo centro de salud privado que forma parte del centro integral(emprendimiento). Se aplicó aprendizaje supervisado en la implementación de algunastécnicas convencionales de machine learning, que se compararon con un modelo de red neuronal artificial para determinar el modelo que pueda obtener la mayor exactitud en el aprendizaje del diagnóstico de diabetes. Como resultado se obtuvo que la red neuronal artificial ANN alcanzó el 99.33% de exactitud frente a las máquinas de soporte vectorial SVM que alcanzaron un 98.6% de exactitud en la clasificación de pacientes con diabetes mellitus, por lo que en vista que el modelo de RNA, es quien aprendió a detectar diabetes con mayor exactitud es el recomendado a usarse en el centro integral de diagnóstico, para la derivación rápida a un centro de salud privado que pueda aplicar un tratamiento rápido y efectivo .

Palabras Claves: Diabetes, Emprendimientos, Centros de Salud Privada, Machine Learning, Aprendizaje Supervisado.

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LEIRD2022.1.1.208
ISBN: 978-628-95207-3-6 ISSN: 2414-6390

I. INTRODUCCION

El *machine learning* (ML) es una de las subramas de la inteligencia artificial (IA) que se ocupa de las formas en que las máquinas aprenden de la experiencia[1]. Sin embargo, algunos informáticos opinan que los términos AI y ML son idénticos debido a la posibilidad de aprender de la experiencia, que es la característica principal de la entidad denominada sistema inteligente[2]. Según la definición detallada del término *Machine Learning* dada por [3] en la que "se dice que un sistema informático ha aprendido de la experiencia E con respecto a alguna tarea T si medido el desempeño D en la tarea en T, mejora con la experiencia E". Durante muchos años, ML ha resuelto muchos problemas sofisticados y problemas complejos del mundo real en las áreas de aplicación como marketing, aplicaciones comerciales, procesamiento de lenguaje natural, atención médica, sistema de vehículos autónomos, robots inteligentes, cambio climático, procesamiento de imágenes, voz, juegos, entre otros. Las técnicas de ML se han utilizado para la predicción y el diagnóstico de muchas enfermedades, como la pandemia de COVID-19[4], la malaria, la fiebre tifoidea,

las enfermedades de las arterias coronarias, la diabetes mellitus[5], entre otras. Los algoritmos de ML generalmente se basan en el enfoque de prueba y error, que es bastante opuesto a los algoritmos convencionales.

Las tareas de ML se clasifican en cuatro categorías amplias, a saber, aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi supervisado y aprendizaje por refuerzo [6]. El aprendizaje supervisado deduce una función a partir de datos de entrenamiento etiquetados, el aprendizaje no supervisado deduce una función de los datos de entrenamiento no etiquetados y el aprendizaje semi supervisado o activo deduce una función eligiendo la muestra más informativa para etiquetar y luego entrenar el modelo[7], mientras que el aprendizaje por refuerzo interactúa con un entorno dinámico en donde se recompensan los comportamientos deseados y penaliza los no deseados[8]. Por el contrario, cuando se utilizan datos etiquetados y no etiquetados para el procesamiento de entrenamiento, este proceso de entrenamiento se denomina semisupervisado[9].

En este estudio, se implementarán algunos de los modelos convencionales de aprendizaje automático *machine learning* (ML) como *Support Vector Machine* (SVM), *K-Nearest Neighbor*(KNN), *Random Forest*(RF), *Naive Bayes*(NB),*Gradient Boosting*(GB),*Logistic Regression* (LG) cuya exactitud se comparara con la de un algoritmo de aprendizaje profundo *deep learning* que usa redes neuronales artificiales *artificial neural network* (ANN) con el objetivo de determinar el algoritmo que mejor aprenda a detectar la diabetes mellitus tipo-2.

La diabetes mellitus (DM), comúnmente conocida como diabetes, es un grupo de trastornos metabólicos del metabolismo de los carbohidratos en los que la glucosa se infrautiliza, lo que produce hiperglucemia (aumento de la concentración de glucosa en la sangre)[10]. Los principales síntomas de la DM incluyen micción frecuente, aumento de la sed y el hambre. Si no se trata, la diabetes puede causar muchas complicaciones, como la cetoacidosis diabética y el coma hiperosmolar no cetósico[11]. Las complicaciones a largo plazo causadas por la DM incluyen enfermedades cardiovasculares, accidentes cerebrovasculares, insuficiencias renales crónicas, úlceras en los pies y daños en los ojos y los nervios. La diabetes ocurre debido a que el páncreas no produce suficiente insulina o las células del cuerpo no responden adecuadamente a la insulina producida[12] . La DM es una de las enfermedades más mortales del mundo, especialmente en los países desarrollados. Sin embargo, se está volviendo más frecuente en los países en vías de desarrollo como Ecuador, al tiempo que representa más amenazas para las personas que no cuentan con cobertura del sistema de salud pública para este tipo de enfermedad[13].

En el Ecuador la diabetes afecta la vida de más de 1.3 millones de ecuatorianos, lo que representa más del 7.5 % de la población[14]. Cuando considera la magnitud de ese número, es fácil entender por qué todos deben estar al tanto de los signos de la enfermedad. La diabetes no tratada puede causar complicaciones graves y, finalmente, convertirse en una amenaza para la vida, pero la detección temprana aumenta la probabilidad de controlar con éxito la enfermedad con un plan de tratamiento eficaz[15].

De hecho, si detecta los signos de un problema potencial en la etapa de prediabetes de la enfermedad, es posible que pueda detener la progresión antes de que desarrolle diabetes tipo-2[16]. Comience por familiarizarse con los factores de riesgo de la diabetes y los signos que debe observar que podrían indicar la aparición de la enfermedad.

Si se tiene diabetes, el cuerpo no produce suficiente insulina o no puede usar la insulina que produce tan bien como debería, cuando no hay suficiente insulina o las células dejan de responder a la insulina, demasiado azúcar en la sangre permanece en el torrente sanguíneo, con el tiempo, eso puede causar problemas de salud graves, como enfermedades cardíacas, pérdida de la visión y enfermedad renal[17]. Todavía no existe una cura para la diabetes, pero perder peso, comer alimentos saludables y mantenerse activo realmente puede ayudar, al igual de tomar los medicamentos según sea necesario, obtener educación y apoyo para el autocontrol de la diabetes y asistir a las citas de atención médica también pueden reducir el impacto de la diabetes en su vida[18].

La diabetes es una condición de salud crónica de larga duración que afecta la forma en que su cuerpo convierte los alimentos en energía, la mayor parte de los alimentos que se consumen se descomponen en azúcar (también llamada glucosa) y se liberan en el torrente sanguíneo, cuando el nivel de azúcar en la sangre sube, le indica al páncreas que libere insulina; la insulina actúa como una llave para permitir que el azúcar en la sangre ingrese a las células de su cuerpo para usarla como energía[19].

II. MATERIALES Y METODOS

A. Dataset

TABLA I
CARACTERÍSTICAS Y ETIQUETA

Variables Independientes	Descripción	Unidad
Embarazos	numero de veces de embarazo	-
Sexo	sexo M(1),F(0)	-
Glucosa	concentración de la glucosa	ml/dl
PresionSanguinea	presion arterial diastólica	mm.Hg
PliegueCutaneo	grosor pliegue cutaneo triceps	mm
Insulina	insulina sérica a 2-horas	µU/ml
IndiceDeMasaCorporal	índice de masa corporal	kg/m^2
PedigriDiabetesFuncion	función pedigrí diabetes	-
Edad	edad de la persona en años	-
Etiqueta		
DiabetesResultado	Si(1),No(0)	-

El un conjunto de datos de diagnóstico de pacientes con DM tipo 2 que fue recopilado de varios centros de salud privados de la ciudad de Guayaquil, en Ecuador. Se conformo un dataset que contiene 2768 registros de pacientes cuyas características son: pregnant_times, glucose, blood_pressure , tst, insulin , bmi , dpf , age , is_diabetic(etiqueta). La Tabla I, muestra la descripción de

unidades y rangos de atributos de riesgo del conjunto de datos. Para la implementación de las técnicas de aprendizaje de maquina supervisado se utilizará Python[20] como herramienta de programación que se ejecuta sobre una máquina virtual de Google llamada Colab[21] que está configurada con todas las librerías requeridas para el uso de machine learning y deep learning, además se necesitará una conexión a internet con un amplio ancho de banda.

B. Revisión de la Literatura

Support Vector Machine

La máquina de vectores de soporte (SVM) es un algoritmo de aprendizaje automático supervisado elegante, potente y uno de los más utilizados. SVM se utiliza tanto para problemas de tareas de aprendizaje automático de regresión como de clasificación debido a su capacidad para predecir patrones separables de forma no lineal mediante la proyección de la función original en un hiperplano (espacio de dimensiones superiores). SVM es un algoritmo no paramétrico que recupera el conjunto de datos de entrenamiento para almacenarlos todos[20]. El algoritmo SVM resuelve problemas de regresión usando funciones lineales, mientras que, en el caso de problemas de regresión no lineal, asigna el conjunto del vector de entrada (a) a un espacio n-dimensional llamado espacio de características (b)[21]. Sin embargo, para datos de entrenamiento multivariante (a_n) se expresa en un número N de observaciones ((b_n), como un conjunto de respuestas observadas. La función lineal se puede representar como:

$$f(x) = x'\beta + b \quad (1)$$

El objetivo es hacerlo lo más ancho posible: $f(x)$ con $\beta'\beta$ como valores mínimos de norma, y como tal el problema encaja en la función de minimización.

$$J(\beta) = \frac{1}{2}\beta'\beta \quad (2)$$

Por lo tanto, con una condición especial de los valores de todos el residual no más de ε, como en la siguiente ecuación:

$$\forall_n: |y_n - (x_n'\beta + b)| \leq \varepsilon \quad (3)$$

K-Nearest Neighbor

K-vecino más cercano (KNN) es uno de los algoritmos de ML supervisado más simples que se basan en la hipótesis "cosas que parecen iguales" [22]. El algoritmo es no paramétrico y es un clasificador supervisado utilizado para las tareas de regresión y clasificación[23]. En ambas tareas, las características de entrada consisten en K ejemplos de entrenamiento más cercanos o el conjunto de datos en la característica espacio, mientras que el algoritmo se basa en datos etiquetados para el proceso de aprendizaje para producir salidas apropiadas para características de entrada no etiquetadas. La idea detrás del algoritmo KNN es que, si una muestra tiene k vecinos más similares en el espacio de características, la mayoría de las muestras pertenecen a una determinada categoría, entonces la muestra también pertenece a esta categoría[24]. El método de votación se utiliza generalmente en la tarea de clasificación, es decir, la etiqueta de categoría que aparece frecuentemente en la muestra k se selecciona como la predicción resultada, mientras que, en el caso de la tarea de regresión, el promedio se utiliza el método donde las etiquetas de salida de valor real de la k muestran se utilizan como resultado de la predicción [25].

Random Forest

El algoritmo del árbol de decisiones aleatorias (RF) fue propuesto en 1995 por bell LABS Ho, que fue defendido agregando muchos clasificadores para mejorar la precisión de la predicción[26]. La idea detrás el algoritmo es combinar múltiples clasificadores de árboles de decisión, como *bagging* y *ramdom space*, para hacer predicciones y obtener el resultado final mediante votos decisorios [27]. La Fig 1 muestra el principio de llenar el bosque aleatorio.

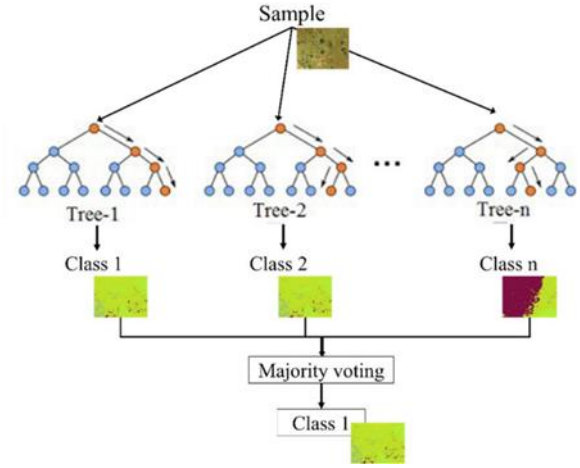

Fig. 1 Random Forest

Naive Bayes

Naive Bayes (NB) también es un algoritmo de aprendizaje automático supervisado, basado en el teorema de Bayes. Aprende estimando la probabilidad previa de cada clase usando un conjunto de datos de entrenamiento [28]. El teorema de Bayes es en la ecuación (4) a continuación:

$$P\left(\frac{A}{B}\right) = \frac{P(\frac{B}{A})(A)}{P(B)} \quad (4)$$

Gradient Boosting

El algoritmo *Gradient Boosting* (GB) o aumento del gradiente combina un conjunto de aprendices débiles para construir un aprendiz fuerte[29]. A diferencia del algoritmo de aprendizaje por impulso o *Boosting*, donde los modelos se hacen de forma independiente[30]. El aumento de gradiente hace que sus modelos sean secuenciales por iteración para minimizar el error de los modelos aprendidos anteriormente:

$$f(x) = \sum_{m=0}^{m} f_m(x) \quad (5)$$

El modelo de conjunto se puede optimizar reduciendo el error de generalización esperada L como se muestra en la ecuación (6):

$$L = \sum_i^n (y_i - \hat{y}_i) * 2 \quad (6)$$

Logistic Regression

El algoritmo *Logistic Regression* (LR) o regresión logística en *machine learning* (ML) es una técnica de regresión adaptativa que construye predicadores como una combinación booleana de covariables binarias[31]. El algoritmo se utiliza para una tarea de clasificación con el objetivo de encontrar una única expresión booleana que prediga un resultado binario. En el caso de una tarea de regresión, muchas expresiones booleanas pueden investigarse e integrarse simultáneamente en un modelo de regresión lineal[32].

Artificial Neural Network – ANN

La neurona artificial según la Fig. 2 es el bloque de construcción central de una red neuronal artificial. La estructura de una neurona artificial es muy similar a una neurona biológica[33], consta de 3 partes principales, peso y sesgo como una dendrita denotada por w y b respectivamente, salida como un axón denotado por y, y función de activación como un cuerpo celular (núcleo) denotado por f(x). La x son las señales de entrada recibidas por la dendrita. En las neuronas artificiales, la entrada y el peso se representan como un vector, mientras que el sesgo se representa como un escalar. La neurona artificial[34] procesa las señales de entrada realizando un producto punto entre el vector de entrada y el vector de peso, agrega el sesgo, luego aplica una función de activación y finalmente propaga el resultado a otras neuronas.

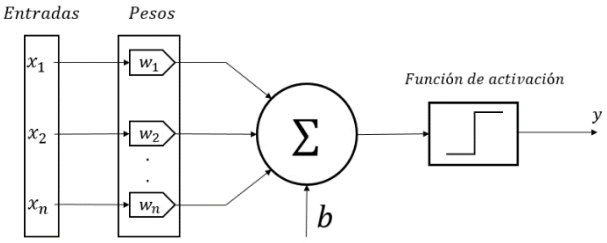

Fig.2 Neurona Artificial

La Red Neuronal Artificial o ANN, es un modelo computacional que imita la forma en que funcionan las células nerviosas en el cerebro humano[35]. Las redes neuronales artificiales (ANN) utilizan algoritmos de aprendizaje que pueden hacer ajustes de forma independiente, o aprender, en cierto sentido, a medida que reciben nuevos datos. Esto las convierte en una herramienta muy eficaz para resolver cualquier modelado de datos estadísticos no lineales[36]. Las ANN de *deep learning* o aprendizaje profundo juegan un papel importante en el aprendizaje automático (ML) y son compatibles con un campo más amplio de la tecnología de inteligencia artificial (IA).

La Red Neuronal Artificial según Fig. 3 tiene tres o más capas que están interconectadas, la primera capa consta de neuronas de entrada las mismas que envían datos a las capas más profundas, que a su vez enviarán los datos de salida finales a la última capa de salida; todas las capas internas están ocultas y están formadas por unidades que cambian adaptativamente la información recibida de una capa a otra a través de una serie de transformaciones[37]. Cada capa actúa como una capa de entrada y salida que permite a la ANN comprender objetos más complejos, en conjunto, estas capas internas se denominan capa neural y las unidades en la capa neuronal intentan aprender sobre la información recopilada al pesarla de acuerdo con el sistema interno de ANN[38]. Estas pautas permiten que las unidades generen un resultado transformado, que luego se proporciona como salida a la siguiente capa.

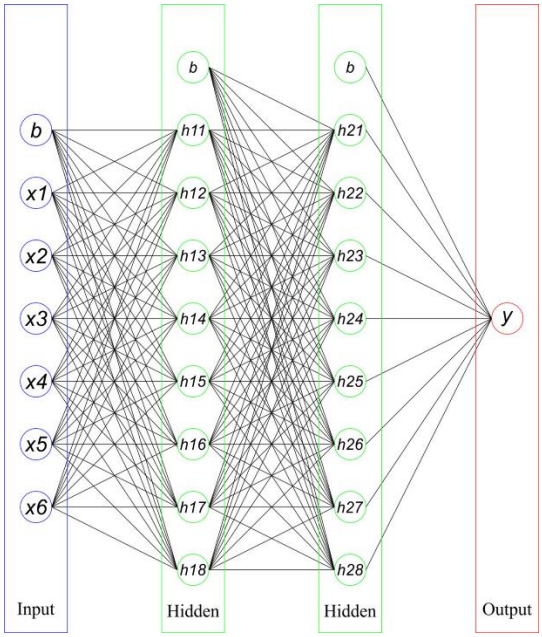

Fig. 3 Red Neuronal Artificial

Métricas de Evaluación

TABLA II
MATRIZ DE CONFUSIÓN

		Predicción 0	Predicción 1
Real	0	TN	FP
	1	FN	TP

Matriz de Confusión. – Según la Tabla II, la matriz de confusión[39] es el elemento sobre el cual se basan todas las métricas de clasificación, en ella se agrupan los valores clasificados por un determinado modelo (0) es *Negative* y (1) es *Positive*. Cuando el modelo clasifica adecuadamente se tienen dos valores. Verdaderos Positivos, cuando el modelo ha predicho que SI y en realidad SI. Verdaderos Negativos, aquí el modelo ha predicho que NO y en realidad es un NO. Cuando no ha clasificado adecuadamente se tienen los siguientes valores. Falso Positivo, cuando el modelo ha predicho que SI y en realidad es un NO. Falso Negativo, es cuando el modelo ha predicho que NO pero en realidad es SI.

En este estudio, se emplean dos técnicas de evaluación para determinar el desempeño de cada modelo de aprendizaje predictivo desarrollado basado en varios algoritmos de ML supervisados. Estas técnicas incluyen lo siguiente:
La precisión se utiliza para evaluar los modelos predictivos de ML supervisados. La precisión tiene la siguiente definición:

1)Accuracy.- *Accuracy* o exactitud es la cantidad de predicciones positivas que fueron correctas.

$$\text{Accuracy} = \frac{(TN+TP)}{(FP+TP)+(TN+FN)} \qquad (7)$$

2) La curva ROC o curva Característica Operativa del Receptor es el gráfico, que expone el rendimiento de un clasificador binario en función del umbral de corte, exponiendo la tasa de verdaderos positivos (TPR) contra la tasa de falsos positivos (FPR)[40] .

III. METODOLOGÍA

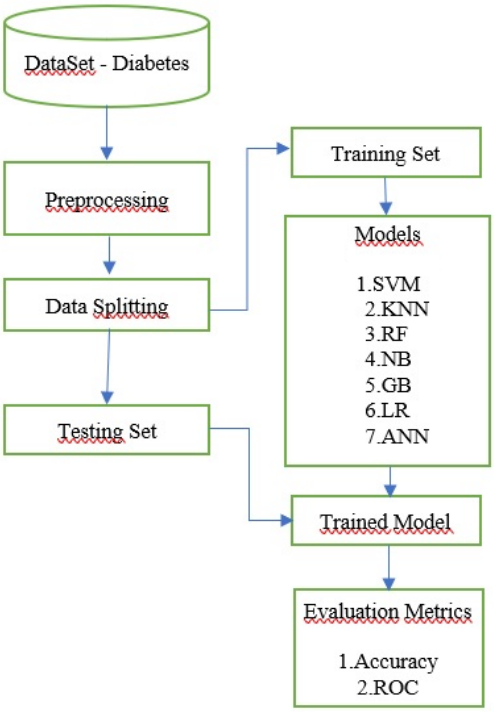

Fig. 4 Modelos de ML

La metodología utilizada está basada en técnicas de aprendizaje automático o *machine learning* al igual que una técnica de aprendizaje profundo o *deep learning* por medio de una red neuronal artificial, los pasos a seguirse se basaron en:

1)Tratamiento de los Datos
2)Creación del Modelo
3)Fase de entrenamiento
4)Fase de Prueba
5)Evaluación del modelo con sus métricas
6)Aplicación del modelo

Fase-I
Se ha trabajado con un dataset que contiene 2768 registros de información de pacientes con diabetes y sin diabetes, la información proviene de varios centros de salud privados de la ciudad de Guayaquil-Ecuador, quienes están interesados en la creación del modelo de *machine learning*.

Del total de registros se designó el 20% para pruebas por lo que X_test tiene 554 registros, el 80% restante se designaron para entrenamiento por lo que X_train tiene 2214 registros.

```
print(X_train.shape,X_test.shape)

(2214, 9) (554, 9)
```

Fig.5 Registros para Train y Test

Los conjuntos de datos fueron escalados según Fig.5 para que no exista diferencias por las magnitudes de cada una de las variables de entrada.

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
# El escalado se ajusta con la media y la
#desviación estandar del conjunto de entrenamiento

scaler.fit(X_train)
# Se aplica la transformación a los datos
X_train_norm = scaler.transform(X_train)
X_test_norm = scaler.transform(X_test)

X_train=X_train_norm
X_test=X_test_norm
```

Fig.6 Escalamiento o Estandarización

Fase II

En esta fase se exponen las etapas que van desde la creación del modelo, entrenamiento, prueba, evaluación y aplicación de estos, por lo que destaca de todos los modelos la red neuronal artificial. El modelo de red neuronal artificial es creado con Keras, tiene una capa de entrada con 9 neuronas y dos capas ocultas más una capa de salida, según (Fig. 7).

```
Model: "sequential_4"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_20 (Dense)            (None, 256)               2560

 dense_21 (Dense)            (None, 512)               131584

 dense_22 (Dense)            (None, 64)                32832

 dense_23 (Dense)            (None, 8)                 520

 dense_24 (Dense)            (None, 1)                 9

=================================================================
Total params: 167,505
Trainable params: 167,505
Non-trainable params: 0
```

Fig.7 Modelo de ANN

La creación de la matriz de confusión Fig. 8 fue muy importante para poder probar las métricas de clasificación, la matriz resultante está relacionada con el mejor de los modelos (ANN).

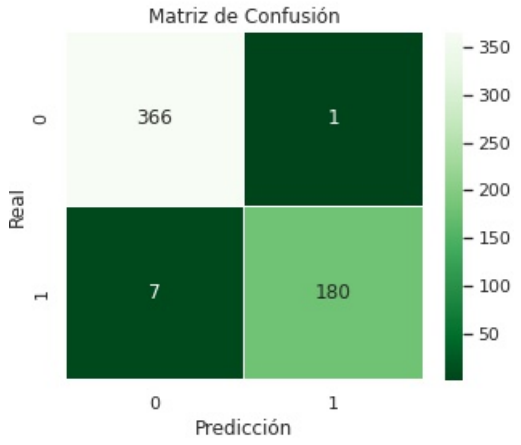

Fig.8 Matriz de Confusión de la ANN

IV. Resultado, Discusion y Conclusion

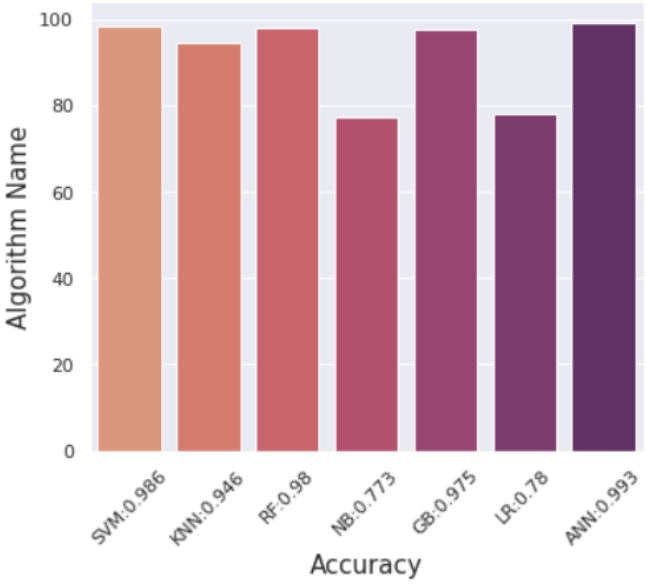

Fig.9 Exactitud de los Modelos

En (Fig. 9) se refleja la exactitud o *accuracy* alcanzado por todos los modelos de *machine learning*, así como el de *deep learning* en el que destaca la red neuronal artificial ANN que refleja una exactitud aproximada del 99.3% y que además se verifica en la curva de la Fig. 10.

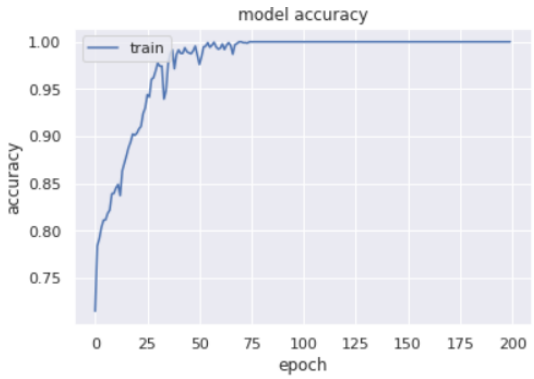

Fig.10 Curva de exactitud del modelo de ANN

La pérdida del modelo según la Fig. 11 a la hora de predecir es aceptable ya que alcanza valores mínimos, corroborando el nivel de exactitud del 99.3% aprox.

Fig.11 Curva de Perdida

Se puede concluir que los modelos de la Fig. 9 han aprendido a realizar adecuadamente la clasificación, en el caso del SVM se obtuvo como resultado una exactitud de predicción del 98.6% frente al modelo basado en ANN con el cual se alcanzó el 99.3% de exactitud y cuyo modelo se refleja en la Fig.12.

```
#Creación del Modelo - ANN
model = Sequential()
model.add(Dense(256,input_dim=9,kernel_initializer='uniform',activation='relu'))
model.add(Dense(512,kernel_initializer='uniform',activation='relu'))
model.add(Dense(64,kernel_initializer='uniform',activation='relu'))
model.add(Dense(8,kernel_initializer='uniform',activation='relu'))
model.add(Dense(1,kernel_initializer='uniform',activation='sigmoid'))

#Compilacion del Modelo
model.compile(loss='binary_crossentropy', optimizer='adam',metrics=['accuracy'])
#Entrenamiento del Modelo
history = model.fit(X_train, y_train, validation_data=(X_test, y_test),
                    epochs=200,batch_size=32, verbose=0)
```

Fig.12 Modelo ANN

El predictor implementado con redes neuronales artificiales, aprendió a predecir adecuadamente mediante clasificación con un elevado índice de exactitud que otros algoritmos, con él se puede determinar con un alto grado de confiabilidad si un paciente tiene diabetes o no; el uso del predictor será de gran ayuda en al sector de la salud porque puede diagnosticar la presencia o no de la diabetes en un paciente en cualquier momento y a cualquier hora, tiene la capacidad para revisar un gran número de exámenes de forma efectiva. Por otra parte, los organismos del sistema de salud podrían brindar un mejor servicio ajustándose a las políticas públicas para agilizar los procesos .de diagnóstico sin que esto afecte al presupuesto del centro de salud que use el modelo creado.

REFERENCIAS

[1] M. Varone, D. Mayer, and A. Melegari, "What is Machine Learning? A definition," *Expert System*, 2020. .

[2] J. Cabanelas Omil, "Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?," *Mercados y Negocios*, no. 40, pp. 5–22, 2019.

[3] A. Panesar, "What Is Machine Learning?," in *Machine Learning and AI for Healthcare*, 2021.

[4] D. P. Pérez, R. S. Bustillos, C. M. Mora, and M. Botto-Tobar, "Prediction of Covid19 with the use of Random Forests Algorithm and Artificial Neural Networks," *Ecuadorian Sci. J.*, vol. 4, no. 2, pp. 101–110, Sep. 2020.

[5] M. E. Baldeón *et al.*, "Prevalence of metabolic syndrome and diabetes mellitus type-2 and their association with intake of dairy and legume in Andean communities of Ecuador," *PLoS One*, vol. 16, no. 7 July, 2021.

[6] J. Luna, "Tipos de aprendizaje automático," *Medium*, 2018.

[7] V. Roman, "Aprendizaje No Supervisado en Machine Learning: Agrupación | by Victor Roman | Ciencia y Datos | Medium," *Medium*, 2019.

[8] Mauricio Arango, "Introducción al Aprendizaje por Refuerzo," *Oracle A-Team*, no. August, 2019.

[9] C. González-García, "En qué consiste el aprendizaje automático (machine learning) y qué está aportando a la Neurociencia Cognitiva," *Cienc. Cogn.*, vol. 12, no. 2, 2018.

[10] R. N. Fatimah, "Diabetes Melitus Tipe 2," *Fak. Kedokt. Univ. Lampung*, vol. 4, 2015.

[11] F. Gómez Rodríguez, P. Ruiz Alcantarilla, L. Rodríguez Félix, A. Martín Santana, and E. Zamora Madaria, "Alteración del receptor Fc(IgG) de los monocitos en la cetoacidosis diabética y el coma hiperosmolar no cetósico.," *Med. Clin. (Barc).*, vol. 84, no. 1, 1985.

[12] J. Vlaški and I. Vorgučin, "Diabetes mellitus type 2 in children and adolescents," *Paediatr. Croat. Suppl.*, vol. 63, 2019.

[13] J. R. Vielma Guevara, J. del C. Villarreal Andrade, and L. V. Gutiérrez Peña, "Pandemia por el SARS-CoV-2: aspectos biológicos, epidemiológicos y clínicos," *Observador del Conocimiento. Revista Especializada de Gestión Social del Conocimiento*, vol. 5, no. 3. 2020.

[14] C. Urrea and A. Mignogna, "Development of an expert system for pre-diagnosis of hypertension, diabetes mellitus type 2 and metabolic syndrome," *Health Informatics J.*, vol. 26, no. 4, 2020.

[15] F. Garmendia-Lorena, "El tratamiento actual de la Diabetes Mellitus Tipo 2," *Diagnóstico*, vol. 59, no. 1, 2020.

[16] X. Liu, S. Wu, Q. Song, and X. Wang, "Reversion from pre–diabetes mellitus to normoglycemia and risk of cardiovascular disease and all-cause mortality in a chinese population: A prospective cohort study," *J. Am. Heart Assoc.*, vol. 10, no. 3, 2021.

[17] E. Menéndez, "TRATAMIENTO NO FARMACOLÓGICO EN LA ENFERMEDAD RENAL POR DIABETES," *Rev. la Soc. Argentina Diabetes*, vol. 51, no. 3, 2018.

[18] A. Y. Forero, J. A. Hernández, S. M. Rodríguez, J. J. Romero, G. E. Morales, and G. Á. Ramírez, "La alimentación para pacientes con diabetes mellitus de tipo 2 en tres hospitales públicos de Cundinamarca, Colombia," *Biomédica*, vol. 38, no. 3, 2018.

[19] E. G. Blanco Naranjo, G. F. Chavarría Campos, and Y. M. Garita Fallas, "Insulinización práctica en la diabetes mellitus tipo 2," *Rev. Medica Sinerg.*, vol. 6, no. 1, 2021.

[20] G. Bentacourt, "Las Máquinas de Soporte Vectorial (SVMs)," *Sci. Tech.*, vol. 1, no. 27, 2005.

[21] J. Rodrigo, "Máquinas de Vector Soporte (Support Vector Machines, SVMs)," *Cienciadedatos*, 2017.

[22] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, 2016.

[23] S. Kang, "K-nearest neighbor learning with graph neural networks," *Mathematics*, vol. 9, no. 8, 2021.

[24] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Appl. Sci.*, vol. 1, no. 12, 2019.

[25] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers-A Tutorial," *ACM Computing Surveys*, vol. 54, no. 6. 2021.

[26] D. Patiño Pérez, R. Silva Bustillos, C. Munive Mora, and M. Botto-Tobar, "Predicción de Covid19 con el uso del Algoritmo Random Forest y Redes Neuronales Artificiales," *Ecuadorian Sci. J.*, vol. 4, no. 2, 2020.

[27] L. Breiman, "Randon Forests," *Machinelearning202.Pbworks.Com*, 1999.

[28] V. Roman, "Algoritmos Naive Bayes: Fundamentos e Implementación," *Ciencia y Datos*, 2019. .

[29] A. V. Konstantinov and L. V. Utkin, "Interpretable machine learning with an ensemble of gradient boosting machines," *Knowledge-Based Syst.*, vol. 222, 2021.

[30] O. Sprangers, S. Schelter, and M. De Rijke, "Probabilistic Gradient Boosting Machines for Large-Scale Probabilistic Regression," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2021.

[31] J. C. Giraldo Mejía and F. A. Vargas Agudelo, “Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales,” *Perspectiv@*, vol. 14, no. 13, 2019.
[32] C. Barrionuevo, J. Ierache, and I. Sattolo, “Reconocimiento de emociones a través de expresiones faciales con el empleo de aprendizaje supervisado aplicando regresión logística,” *XXVI Congr. Argentino Ciencias la Comput.*, pp. 491–500, 2020.
[33] J. Archila Rodriguez, “La neurona,” *LA Neurona*, 2017.
[34] J. A.-T. Barrera, “Redes Neuronales Artificiales,” *Univ. Guadalajara*, p. 276, 2016.
[35] B. Martín del Brío and C. Serrano Cinca, “Fundamentos de redes neuronales artificiales: hardware y software,” *Scire Represent. y Organ. del Conoc.*, 1995.
[36] A. Novales, “Estimación de modelos No Lineales,” *Dep. Econ. Cuantitativa Univ. Complut.*, 2016.
[37] O. Agasi, J. Anderson, A. Cole, M. Berthold, M. Cox, and D. Dimov, “What is an Artificial Neural Network (ANN)? - Definition from Techopedia,” *Techopedia*. 2018.
[38] D. Patiño Perez, R. Silva Bustillos, M. Botto-Tobar, and C. Munive Mora, “Análisis de Imágenes de Rayos X por Medio de Redes Neuronales Artificiales,” *Ecuadorian Sci. J.*, vol. 5, no. 1, 2021.
[39] P. Recuero de los Santos, “Machine Learning a tu alcance: La matriz de confusión - Think Big Empresas,” *Luca Telefonica Data Unit*, 2018. .
[40] A. A. Osi *et al.*, “A classification approach for predicting COVID-19 Patient’s survival outcome with machine learning techniques,” *medRxiv*, 2020.