



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES  
UNIDAD MORELIA

MANUAL DE PRÁCTICAS PARA LA MEJORA DE LA  
ENSEÑANZA DEL APRENDIZAJE PROFUNDO  
APLICADO A LA CIENCIA DE DATOS A GRAN  
ESCALA

## INFORME FINAL

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN TECNOLOGÍAS PARA  
LA INFORMACIÓN EN CIENCIAS

P R E S E N T A

BRUCE HIRAM GINORI RODRÍGUEZ

TUTOR

DR. SERGIO ROGELIO TINOCO MARTÍNEZ



ESCUELA  
NACIONAL  
de ESTUDIOS  
SUPERIORES  
**mm**  
UNIDAD MORELIA

CO-TUTOR

DR. HEBERTO FERREIRA MEDINA  
MORELIA, MICHOACÁN FEBRERO 2023

# Agradecimientos institucionales

Agradezco a la licenciatura en Tecnologías para la Información en Ciencias (LTICs), todas las facilidades otorgadas para el desarrollo de este manual, a los estudiantes por su apoyo al contestar la encuesta y a los profesores por su ayuda para realizar este documento.

Agradezco a la DGAPA, UNAM, su apoyo para desarrollar este trabajo que es parte del proyecto PAPIME PE106021 titulado "Propuesta de mejora a la enseñanza del aprendizaje automático aplicado a la Ciencia de Datos a gran escala; dirigido a académicos, académicas y estudiantes de la licenciatura en Tecnologías para la Información en Ciencias (LTICs) en la ENES Morelia"

Agradezco también al comité tutorial sus observaciones y revisiones para mejorar la calidad de este trabajo.

Agradezco a mi asesor Dr. Sergio Rogelio Tinoco Martinez toda su ayuda y apoyo para la terminación de este documento.

Agradezco a los profesores de la LTIC por sus enseñanzas y apoyo durante mi carrera.

# Agradecimientos personales

- A la familia
- A los amigos
- Alguien importante
- Un amigo en especial

# Resumen

Este trabajo forma parte del proyecto PAPIME titulado “Propuesta de mejora a la enseñanza del aprendizaje automático aplicado a la Ciencia de Datos a gran escala” con el número de proyecto PE106021. Se propone la elaboración de un manual de prácticas a nivel licenciatura para las asignaturas en el área de aprendizaje profundo o Deep Learning (DL) usando herramientas a gran escala (Big Data). Con el fin de ayudar a los estudiantes a entender el DL y sus herramientas, utilizando prácticas, que pueden desarrollarse a partir del sexto semestre de la LTICs y de otras licenciaturas afines, en la ENES Morelia que requieren los conocimientos del DL.

La realización de este manual de prácticas, así como el contenido del mismo, fue determinado, modificado y adaptado mediante una encuesta aplicada a los alumnos pertenecientes a carreras relacionadas con el campo de estudio, este instrumento ayudó a determinar las áreas de interés con respecto al DL.

Nuestra población de muestra además de los estudiantes contó con la participación de los docentes y académicos de las diferentes licenciaturas y postgrados dentro de la ENES Morelia y del Campus de la UNAM. La finalidad es obtener información con relación a las materias de la carrera de LTICs, con el fin de adaptar el contenido de estas, para un mejor entendimiento de la inteligencia artificial y el DL.

El manual tiene el propósito de brindar una mejor formación académica a los estudiantes de LTICs y licenciaturas afines al DL. Además, resulta ser una herramienta para que los docentes conocieran las necesidades estudiantiles y coadyuvar al autoaprendizaje en esta área.

El manual se conforma de seis prácticas base y cuatro complementarias que ayudan al entendimiento de la aplicación del DL y el Big Data.

Las prácticas realizadas a lo largo de este proyecto (como libretas de JupyterLab) se encuentran en el siguiente repositorio en la plataforma GitHub:

<https://github.com/to-be-changed/Practicas-DL>

# Abstract

This work is part of the PAPIME project entitled "Proposal to improve the teaching of machine learning applied to large-scale Data Science" with the project number PE106021. This paper proposes the development of a practice manual at the undergraduate level for subjects in the area of deep learning or Deep Learning (DL) on a large scale (Big Data). This in order to implement practices aimed at students from the sixth semester of the Bachelor's Degree in Information Technology in Sciences or other ENES Morelia degrees that have basic knowledge of DL.

The realization of this manual of practices, as well as the content of the same, will be determined, modified and adapted by means of a survey applied to the students belonging to careers related to the field of study, where the areas of interest with respect to the DL can be determined. .

Our main sample population will be the students and/or teachers of the different undergraduate and postgraduate degrees within the ENES Morelia closely related to the LTICs career, this in order to be able to adapt the study plan of the same, including the topics of greatest interest. relevance to machine learning subjects.

This manual has the purpose of providing a better academic training to students of LTICs and degrees related to DL. In addition, it turns out to be a tool for teachers to know the needs of students and contribute to self-learning in this area.

The practices carried out throughout this project ( as JupyterLab notebooks format) are found in the following repository on the GitHub platform:

<https://github.com/to-be-changed/Practicas-DL>

# Índice general

<b>Agradecimientos institucionales</b>	<b>I</b>
<b>Agradecimientos personales</b>	<b>II</b>
<b>Resumen</b>	<b>III</b>
<b>Abstract</b>	<b>V</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Justificación . . . . .	2
1.2. Hipótesis . . . . .	2
1.3. Objetivo . . . . .	3
1.3.1. Objetivo General . . . . .	3
1.3.2. Objetivos Particulares . . . . .	3
1.4. Descripción general . . . . .	5
<b>2. Antecedentes</b>	<b>6</b>
2.1. Aprendizaje Automático. . . . .	7
2.1.1. Inteligencia artificial y aprendizaje automático. . . . .	7
2.1.2. Aprendizaje supervisado . . . . .	7
2.1.3. Aprendizaje no supervisado . . . . .	7
2.1.4. Aprendizaje por refuerzo . . . . .	8
2.1.5. Problemas del aprendizaje automático: ajuste bajo y sobreajuste	8
2.1.6. Aprendizaje profundo . . . . .	8

2.2. Diseño de la Encuesta . . . . .	8
2.3. Ejes y preguntas . . . . .	10
<b>3. Aprendizaje Profundo</b>	<b>17</b>
3.1. Inspiración biológica y aproximadores universales . . . . .	17
3.2. Unidades de procesamiento (neuronas), capas y arquitectura . . . . .	18
3.2.1. Neuronas . . . . .	18
3.2.2. Capas . . . . .	18
3.2.3. Arquitectura . . . . .	19
3.3. Aprendizaje superficial y profundo . . . . .	20
3.4. Componentes importantes . . . . .	21
3.4.1. Capas lineales . . . . .	21
3.4.2. Funciones de activación . . . . .	22
3.4.3. Funciones de pérdida . . . . .	23
3.4.4. Capas convolucionales . . . . .	23
3.4.5. Capas de pooling . . . . .	23
3.4.6. Capas Recurrentes . . . . .	23
3.4.7. Normalización por lotes (Batch Normalization) . . . . .	23
3.4.8. Dropout . . . . .	23
3.4.9. Entrenamiento y métodos de optimización . . . . .	23
3.4.10. Métricas . . . . .	23
3.4.11. Conjuntos de entrenamiento . . . . .	23
3.4.12. Preprocesamiento de datos . . . . .	23
3.5. Transferencia de aprendizaje (Transfer Learning) y ajuste fino (Fine Tuning) . . . . .	23
<b>4. Algoritmos de Deep Learning</b>	<b>24</b>
4.1. Redes Neuronales Convolucionales . . . . .	24
4.2. Data Augmentation . . . . .	25
4.3. Redes Neuronales Recurrentes . . . . .	26
4.4. Transferencia de Aprendizaje . . . . .	26



4.5. TensorFlow . . . . .	26
<b>5. Métricas para Evaluar Modelos</b>	<b>27</b>
5.1. Exactitud . . . . .	27
5.2. Precisión . . . . .	28
5.3. Sensibilidad (Recall) . . . . .	29
5.4. Matriz de Confusión . . . . .	29
<b>6. Metodología de las Prácticas</b>	<b>31</b>
<b>7. Resultados y Discusión</b>	<b>33</b>
<b>Apéndices</b>	<b>34</b>
.1. Prácticas . . . . .	35
.2. Datos . . . . .	35

# Índice de figuras

1.1. Mapa mental del Proyecto. Fuente: Tesis de Licenciatura "Blabla- bla"(Cita o referencia) / Elaboración Propia / Fuente: Nombre de la pagina (URL) . . . . .	4
2.1. Matriz de correlación de los resultados obtenidos en la encuesta, donde los encabezados son las preguntas realizadas en la encuesta ver Tabla 2.2. CITAR EL ARTÍCULO DONDE VIENENN ESTOS GRÁFICOS	12
2.2. Nivel de importancia de ejes de acuerdo a la población encuestada. .	13
2.3. Nivel de importancia de temas de acuerdo a la población encuestada.	14
2.4. Nivel de importancia de herramientas de acuerdo a la población en- cuestada. . . . .	15

# Capítulo 1

## Introducción

En la Escuela Nacional de Estudios Superiores Unidad Morelia (ENES Morelia), perteneciente a la Universidad Nacional Autónoma de México (UNAM) se imparte la Licenciatura en Tecnologías para la Información en Ciencias (LTICs). La carrera cuenta con tres áreas de profundización, siendo el área de Ciencias de la Información la clave en este estudio. Un conjunto de materias de esta área se centra de abarcar el análisis y procesamiento de grandes volúmenes de datos, una actividad que en los años recientes se ha vuelto una herramienta poderosa y crucial para la Ciencia de Datos.

Dado esto, la inclusión de estas materias en el plan curricular de la carrera de LTICs es la decisión más óptima, pero que sufre un retraso gracias a las dificultades de aprendizaje de los temas y la complejidad de los mismo; se requieren conocimientos base de aplicaciones de Inteligencia Artificial (IA), Aprendizaje Automático (Machine Learning) y aprendizaje profundo (Deep Learning)

El DL es un subconjunto del ML ya que automatizan procesos reduciendo la intervención humana gracias al uso de la IA en diferentes servicios y aplicaciones como la realización de tareas analíticas.

En el sector de salud se ha mejorado mucho gracias a la digitalización de sus registros médicos puesto que estos pueden ser engastados por aplicaciones de reconocimiento de imágenes, ayudando al analizar y evaluar más imágenes en menos tiempo ayudando a especialistas médicos.

Estos mismos algoritmos pueden ser usados para la identificación de patrones peligrosos que lleven a fines fraudulentos. El reconocimiento de voz, visión por computadora y otras aplicaciones pueden extraer evidencias y analizarlas de manera rápida y precisa para las fuerzas del orden público.

El propósito de este trabajo es brindar un manual que sirva como apoyo, dando ejemplos prácticos de la vida real y herramientas para la aplicación a más áreas relacionadas al DL y el Big Data.

## 1.1. Justificación

Al probar que es una herramienta poderosa y de gran importancia en la actualidad, en la LTICs se imparten materias donde se enseñan temas de Deep Learning aplicados al Big Data. Estas materias recurren a literatura clásica del área para poder tener un contexto más aterrizado lo que les da un acercamiento teórico que resulta excesivo y abrumador para el alumnado, causando un rezago notorio en este llegando a la deserción de los mismos.

En el entendimiento de que estos materiales son cruciales para el trabajo y desarrollo, este manual busca brindar un apoyo complementario y práctico con ejemplos de la vida real para poder agilizar el desarrollo de las materias y apoyar al alumno con herramientas adicionales.

## 1.2. Hipótesis

La Licenciatura en Ciencias en Tecnologías de la Información (ENES Morelia, UNAM) cuenta con un número significativo de estudiantes con bajo rendimiento en materias de DL debido a la complejidad de las materias y la poca aplicación de ejemplos prácticos. Con la intervención y aplicación de el manual de prácticas, se espera que su desempeño mejore sustancialmente.

## 1.3. Objetivo

### 1.3.1. Objetivo General

Desarrollar un manual de prácticas para la enseñanza del Deep Learning, aplicado a volúmenes de datos a gran escala, dirigido a estudiantes a partir de sexto semestre de la Licenciatura en Tecnologías para la Información en Ciencias o de otras licenciaturas de la ENES Morelia que cuenten con los conocimientos básicos del Deep Learning

### 1.3.2. Objetivos Particulares

1. Desarrollar una encuesta para diagnosticar los conocimientos previos e intereses del alumnado.
2. Establecer los temas que se van a cubrir en el manual de prácticas, relacionados al DL y con fundamento en la encuesta aplicada.
3. Elaborar el marco teórico del manual de prácticas con base en los temas seleccionados.
4. Determinar los ejemplos prácticos del ML que se abordarán en el manual de prácticas, conciliando con los docentes de la LTICs de las asignaturas del área del DL, la pertinencia de las prácticas propuestas enfocadas al Big Data.
5. Implementar los ejemplos prácticos usando el lenguaje Python.
6. Realizar una prueba piloto del manual de prácticas.
7. Realizar el diagnóstico de los resultados de la intervención a través de una encuesta de salida.
8. Publicar los resultados en la página web del proyecto.

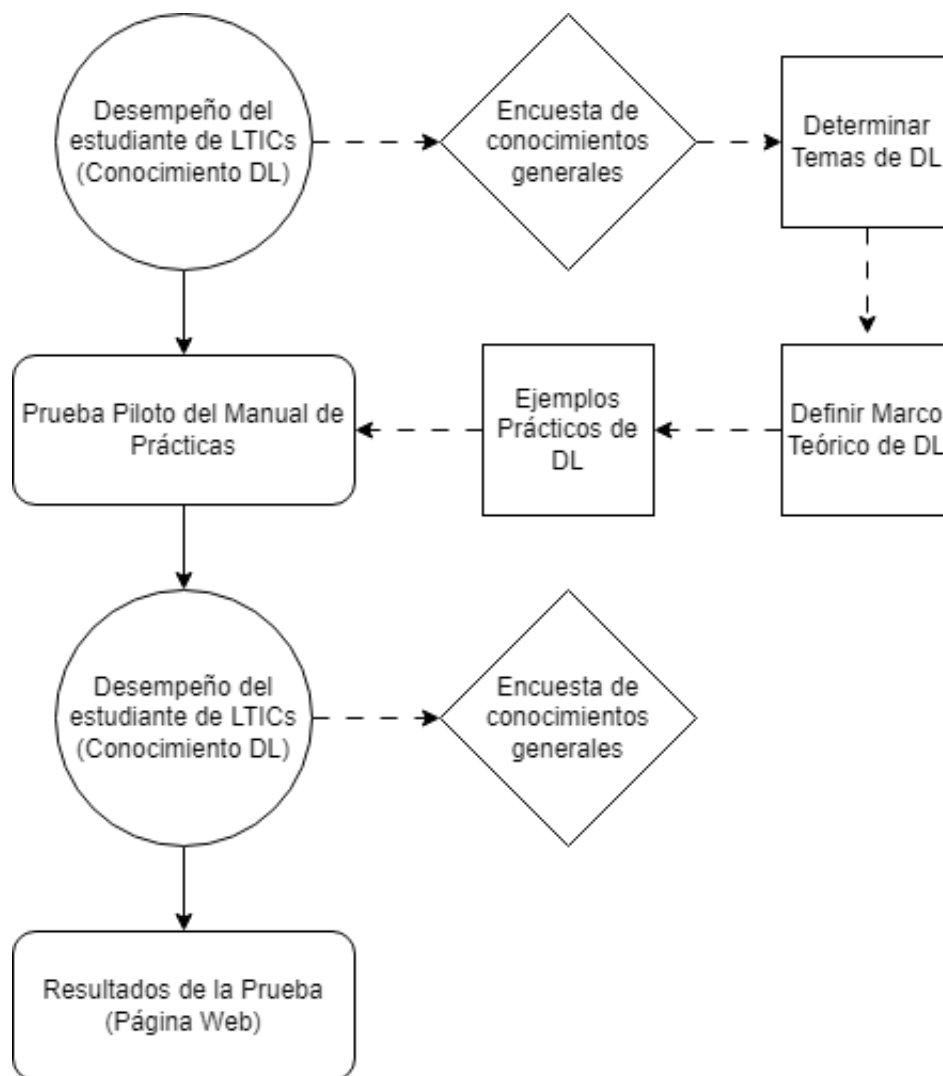


Figura 1.1: Mapa mental del Proyecto. Fuente: Tesis de Licenciatura "Blablabla"(Cita o referencia) / Elaboración Propia / Fuente: Nombre de la pagina (URL)

## 1.4. Descripción general

Este documento está organizado de la siguiente manera:

- El Capítulo 2 se abarcan los antecedentes con respecto a las investigaciones importantes que llevaron a cabo
- El Capítulo 3
- El Capítulo 4
- El Capítulo 5
- El Capítulo 6 El Capítulo 5 enlista y da un resumen de la metodología de las prácticas. En éste se explica como se lleva a cabo la estructura de cada práctica, los datos que se van a usar, la métrica de evaluación
- Finalmente, en el Capítulo 7 se recopilan los datos recabados después de impartir el segundo módulo del diplomado centrado en el DL. Se presentan los resultados relacionados con la mejora del aprendizaje en DL en estudiantes de la Licenciatura en Tecnologías de la Información y Comunicación (LTICs) y académicos con formación afín a esta disciplina.

# Capítulo 2

## Antecedentes

Según LeCun y cols. (2015) el Deep Learning (DL) es un campo de estudio de la Inteligencia Artificial que ha experimentado un crecimiento significativo en las últimas décadas. McCulloch y Pitts (1943) Ubican que sus orígenes se remontan a los primeros intentos de imitar el funcionamiento del cerebro humano y desarrollar modelos computacionales capaces de aprender y realizar tareas complejas.

En su artículo Bengio (2009) señalan que, en sus inicios, el DL enfrentaba limitantes importantes debido a la falta de recursos computacionales y a la falta de conjuntos de datos lo suficientemente grandes y etiquetados. Esto dificultaba el entrenamiento de modelos profundos y limitaba su aplicación en problemas del mundo real.

Con el avance de la tecnología y el aumento en la disponibilidad de datos, el DL ha experimentado una evolución notable. Se han propuesto y desarrollado nuevas arquitecturas y algoritmos como las Redes Neuronales Convolucionales (CNN) y las Redes Neuronales Recurrentes (RNN) que aplicaron Krizhevsky y cols. (2012) y demostraron un rendimiento sobresaliente en tareas como el reconocimiento de imágenes y el procesamiento del lenguaje natural.

En el libro de Goodfellow y cols. (2016) señalan que los descubrimientos en el campo del DL han sido impulsados por la investigación y la experimentación constante. Investigadores de renombre han contribuido con sus trabajos, como Geoffrey Hinton, Yann LeCun y Yoshua Bengio, quienes han desarrollado conceptos clave y han propuesto modelos y algoritmos fundamentales en el campo.



En la actualidad, el Aprendizaje Profundo se encuentra en un estado de continua expansión y aplicación. Ha demostrado resultados prometedores en una amplia gama de aplicaciones, incluyendo reconocimiento de voz, traducción automática, conducción autónoma y diagnóstico médicos como los mostrados por Esteva y cols. (2017)

## **2.1. Aprendizaje Automático.**

El Aprendizaje Automático o Machine Learning (ML) es, según Alpaydin (2010) una disciplina que se enfoca en el desarrollo de algoritmos y técnicas que permiten a los sistemas informáticos aprender patrones y tomar decisiones a partir de los datos sin ser explícitamente programados para realizar tareas específicas

### **2.1.1. Inteligencia artificial y aprendizaje automático.**

Murphy (2012) indican en su libro que el ML es una sub-rama de la Inteligencia Artificial (IA) que se enfoca en el desarrollo de algoritmos y modelos que permiten a las máquinas aprender patrones a partir de los datos y mejorar su desempeño en tareas específicas sin ser programadas explícitamente para cada una de ellas. En el ML, los modelos se ajustan y mejoran a medida que se les proporciona más información, lo que les permite tomar decisiones o realizar predicciones de manera más precisa.

### **2.1.2. Aprendizaje supervisado**

El aprendizaje supervisado es según Hastie y cols. (2009) una técnica de ML en la que se utilizan datos etiquetados para entrenar un modelo de predicción. En este tipo de aprendizaje, se tiene un conjunto de datos de entrenamiento que consiste en pares entrada-salida.

### **2.1.3. Aprendizaje no supervisado**

El aprendizaje no supervisado es según Murphy (2012) una técnica de ML en la que se utilizan datos no etiquetados para entrenar un modelo de agrupamiento o

reducción de dimensionalidad. En este tipo de aprendizaje, el objetivo es encontrar patrones en los datos sin conocer la respuesta correcta.

#### **2.1.4. Aprendizaje por refuerzo**

El aprendizaje por refuerzo es una según Shalev-Shwartz y Ben-David (2014) técnica de ML en la que un agente aprende a través de la interacción con un entorno. En este tipo de aprendizaje, el agente realiza acciones y recibe recompensas o castigos según su desempeño.

#### **2.1.5. Problemas del aprendizaje automático: ajuste bajo y sobreajuste**

Uno de los principales problemas del ML es el ajuste bajo y el sobreajuste. Según Ng (2017) el ajuste bajo ocurre cuando un modelo es demasiado simple y no es capaz de capturar toda la complejidad de los datos. El sobreajuste ocurre cuando un modelo es demasiado complejo y se ajusta demasiado a los datos de entrenamiento, lo que puede resultar en un mal desempeño en nuevos datos .

#### **2.1.6. Aprendizaje profundo**

El D L es una según Goodfellow y cols. (2016) subrama del AA que utiliza redes neuronales profundas para aprender patrones a partir de los datos. Las redes neuronales profundas se componen de múltiples capas de neuronas y son capaces de aprender representaciones jerárquicas de los datos.

### **2.2. Diseño de la Encuesta**

Esta investigación se caracterizó por ser un estudio de tipo:

1. Exploratorio
2. Descriptivo

## 3. Correlacional

## 4. Pre - experimental

con la finalidad de contar con un estudio de caso a través de una sola medición. Para ello se generó una encuesta que fue aplicada a una población conformada por profesores, estudiantes e investigadores de la Universidad Nacional Autónoma de México campus Morelia, Michoacán.

La encuesta fue creada usando una plataforma en linea encuesta.com, (2021) y distribuida por correo electrónico (Gmail principalmente) y redes sociales a personas seleccionadas al azar entre la población de estudio de la universidad mencionada anteriormente. En primer lugar, la muestra se calculó mediante el método de población finita con base a 600 personas. Esta muestra tiene un intervalo de confianza del 95 % y un margen de error del 10 %, como se muestra en la Tabla 2.1

Descripción	Valor
Tamaño de la población	600
Nivel de confianza	95 %
Margen de error	10 %
Tamaño de la muestra	83

Tabla 2.1: Muestra de la población.

La encuesta consistió en ocho preguntas generadas a partir de una revisión crítica de la literatura relacionada con la ciencia de datos de ML, DL y Big Data. Se consultaron a expertos y expertas en el área de ML, cuatro maestros titulares de la ENES Morelia, tres del Tecnológico Nacional de México campus Morelia, dos del Instituto de Investigaciones en Ecosistemas y Sustentabilidad (IIES, UNAM) y uno del Instituto de Materiales UNAM

Ellas y ellos validaron las preguntas seleccionadas, mencionando que se deberían considerar los siguientes puntos:

- Dar diferentes opciones para responder.

- Poner respuestas dicotómicas.
- Usar escala de Likert.

## 2.3. Ejes y preguntas

Gracias a las observaciones, la encuesta se perfeccionó dividiéndola en cuatro ejes principales: *Eje I: Machine Learning; Eje II: Deep Learning; Eje III: Big Data; y Eje IV. Herramientas.*

La escala Likert aplicada para los primeros tres ejes fue: Muy importante (*VImp - Very Important*), Importante (*Imp - Important*), Neutral, Poco importante (*LImp - Less Important*), Nada importante (*NImp - Nothing Important*), Lo desconozco (*dnK - Do not Know*); ver Tabla 2.2.

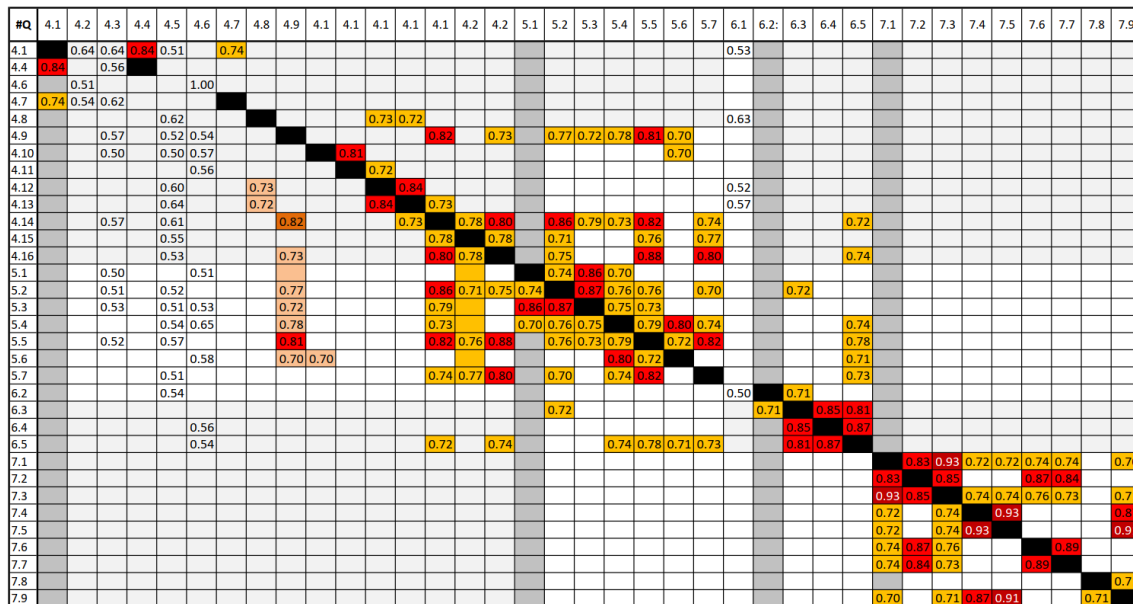
Mientras que para el cuarto eje se usó el nivel de habilidad que se tiene para diferentes herramientas de programación *bajo, medio, alto*.

Como se puede observar, cada escala se abrevió por sus siglas en inglés para que se pueda entender mejor.

#	Descripción de la pregunta	Eje	Tipo de pregunta
Q1	N° de cuenta o empleado UNAM, tu nombre si no cuentas con ellos	-	Opción
Q2	Licenciatura que estudia; 2.1 Ciencias, 2.2 Agroforestales, 2.3 Ciencias Ambientales, 2.4 Ciencia de Materiales Sustentables, 2.5 Ecología, 2.6 Estudios Sociales y Gestión Local, 2.7 Geociencias, 2.8 Geohistoria, 2.9 Tecnologías para la Información en Ciencias, 2.10 Otro (Other)	-	Opción
Q3	Semestre que cursas (1-10), no aplica para profesores	-	Número
Q4	Consideras que los siguientes temas relacionados con el aprendizaje automático (Machine Learning) son: 4.1 Data Science, 4.2 Web Scraping, 4.3 Data Wrangling, 4.4 Machine Learning, 4.5 Data Mining, 4.6 Ensemble Learning, 4.7 Data visualization, 4.8 ML: supervised/unsupervised, 4.9 Binary and multiclass classification, 4.10 EDA, 4.11 Clustering, 4.12 ML model, 4.13 ML evaluation: underfitting, overfitting, 4.14 Cross validation, 4.15 Hyperparameters, regularization, feature engineering, 4.16 PCA	I	Opciones Likert
Q5	Consideras que los siguientes temas relacionados con el aprendizaje profundo (Deep Learning) son: 5.1 NN Shallow & Deep, 5.3 CNN, 5.3 RNN, 5.4 Transfer Learning & Fine-Tuning, 5.5 Dropout, 5.6 Data Augmentation, 5.7 Batch Normalization	II	Opciones Likert
Q6	Consideras que los siguientes temas relacionados con los macrodatos (Big Data) son: 6.1 Concepto, 6.2 Escalado de modelos, 6.3 Analítica a gran escala, 6.4 Sistema de archivos distribuidos, 6.5 Map-Reduce	III	Opciones Likert
Q7	La habilidad que tienes en el manejo de las siguientes herramientas es: 7.1 Tensorflow, 7.2 Spark, 7.3 Keras, 7.4 Fast.ai, 7.5 PyTorch, 7.6 HDFS, 7.7 Kafka, 7.8 Python, 7.9 Scikit-Learn	VI	Opciones Likert
Q8	Consideras importantes incluir algunos temas adicionales relacionados con el ML, DL y el Big Data, no mencionados anteriormene:	-	Abierta

Tabla 2.2: Tabla 2.2: Encuesta aplicada.

También se aplicó un estudio de correlaciones utilizando la bivariada de Pearson y seleccionando únicamente las correlaciones obtenidas en los niveles alto y muy alto [0.7, 0.93], que se muestran en la Figura 2.X.



De acuerdo al análisis de correlaciones, se identificaron áreas de oportunidad según porcentaje de importancia que respondieron los encuestados. En la Figura 2.x se muestra esta importancia.

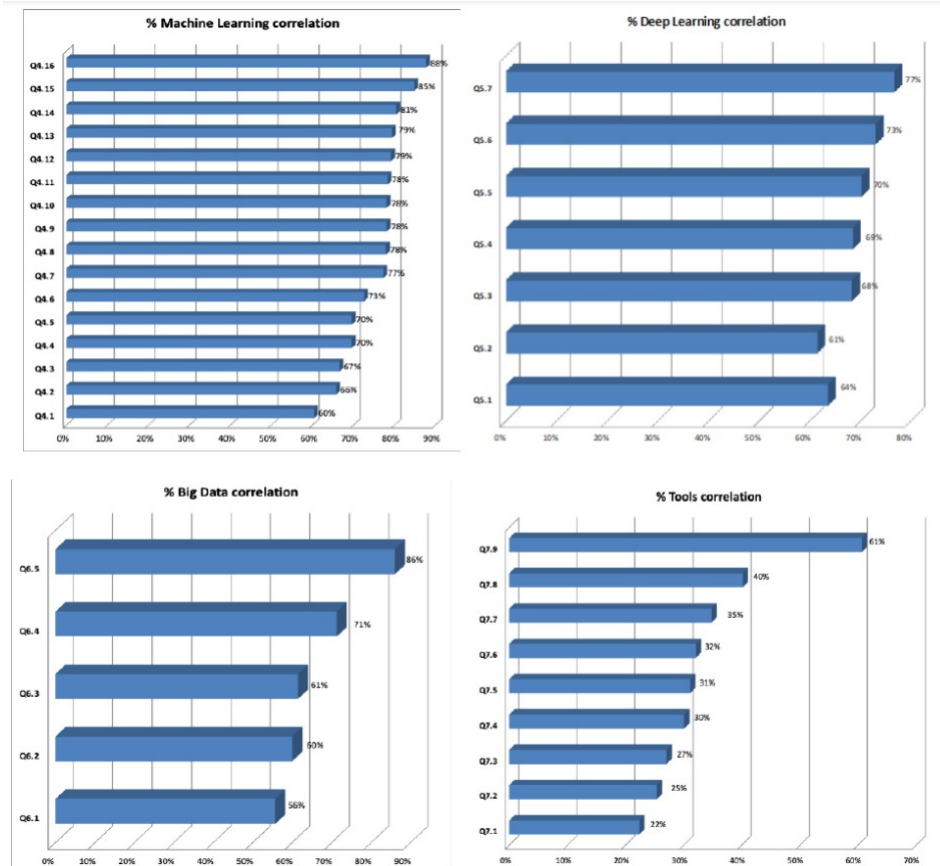


Figura 2.2: Nivel de importancia de ejes de acuerdo a la población encuestada.

De acuerdo con la encuesta desarrollada, se observó que los encuestados tenían cierto desconocimiento en algunos temas. En las Figuras 2.10 y 2.11 los temas se muestran por eje, ordenados por nivel de desconocimiento: No sé (dnK), Nada importante (NImp), Menos importante (LImp), Neutro Importante (Imp), Muy importante (VImp) y para las herramientas, la escala es: No sé (dnK), Corta, Media y Alta.



Figura 2.3: Nivel de importancia de temas de acuerdo a la población encuestada.





Figura 2.4: Nivel de importancia de herramientas de acuerdo a la población encuestada.

Como se observa en las Figuras, los temas que más desconocen las y los participantes (color verde) son los temas a desarrollar en el manual para obtener un mayor impacto académico, así mismo se tomó en cuenta la importancia que la población encuestada le dió a los temas, resultando en los siguientes temas.

1. Clasificación Binaria
2. Redes Convolucionales
3. Data Augmentation
4. Redes Neuronales Recurrentes
5. Transferencia de Aprendizaje
6. TensorFlow Distribuido

De acuerdo a estos resultados se propone mejorar el aprendizaje con prácticas orientadas a reforzar estos temas. En la Tabla 5.1 y la Tabla 5.2 se muestran las prácticas finalmente propuestas para mejorar en los temas que observamos como área de oportunidad. Cabe resaltar que se observó que las y los encuestadas (os) prefieren orientar la intervención hacia la aplicación práctica de los conocimientos.

En el siguiente capítulo se describen brevemente los temas correspondientes al DL determinado con fundamento en la encuesta aplicada, antes de proponer el esquema de las prácticas que conforman el manual.

Debido a la demanda del conocimiento y manejo del DL es necesario tener una lista de temas a considerar para definir el número y la cantidad de prácticas a realizar. A continuación se muestran los principales temas a considerar.

# Capítulo 3

## Aprendizaje Profundo

### 3.1. Inspiración biológica y aproximadores universales

Como señalan LeCun y cols. (2015) la inspiración biológica es un concepto fundamental en el desarrollo del Aprendizaje Profundo (DL), el cual se basa en la estructura y funcionamiento del cerebro humano para diseñar redes neuronales artificiales. Esta aproximación se deriva de la idea de que el cerebro es capaz de realizar tareas cognitivas complejas, como el reconocimiento de patrones, la toma de decisiones y el procesamiento del lenguaje, de una manera altamente eficiente y adaptable.

El enfoque biológico busca imitar la estructura y el comportamiento de las redes neuronales del cerebro, que están compuestas por un gran número de neuronas interconectadas. Las redes neuronales artificiales se construyen utilizando unidades básicas llamadas Neuronas Artificiales, que están organizadas en capas y se conectan entre sí mediante pesos sinápticos. Cada neurona artificial procesa las señales de entrada, aplica una función de activación y transmite la señal a las neuronas de la capa siguiente.

Este enfoque biológico ha sido ampliamente explorado y ha demostrado ser efectivo en la solución de diversos problemas de aprendizaje automático, como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la conducción autónoma

como es aplicado por Serre (2018). Al modelar las redes neuronales de manera similar al cerebro, se espera capturar algunas de las capacidades cognitivas y de aprendizaje del sistema biológico.

## 3.2. Unidades de procesamiento (neuronas), capas y arquitectura

### 3.2.1. Neuronas

LeCun y cols. (2015) que las neuronas o unidades de procesamiento son elementos fundamentales que imitan el funcionamiento de las neuronas biológicas del cerebro humano. Estas unidades son la base de las redes neuronales artificiales y se organizan en capas para realizar tareas de aprendizaje automático.

Las neuronas artificiales reciben señales de entrada, que son ponderadas por unos pesos sinápticos. Luego, estas señales ponderadas se procesan en la neurona mediante una función de activación, que determina si la neurona debe activarse o no. La salida de la neurona se transmite a otras neuronas en la capa siguiente o se considera como la salida final del modelo.

Este enfoque está inspirado en la biología, ya que se basa en cómo las neuronas biológicas procesan y transmiten información mediante conexiones sinápticas.

### 3.2.2. Capas

En el DL, una capa se refiere a una colección de neuronas o unidades de procesamiento que realizan un tipo específico de cálculo en los datos de entrada. Una red neuronal profunda generalmente consta de múltiples capas, cada una de las cuales transforma los datos de entrada en una representación más abstracta que captura características y patrones de nivel superior.

Hay varios tipos de capas comúnmente utilizadas en el aprendizaje profundo, que incluyen:

1. Capa de entrada: esta capa recibe los datos de entrada y los pasa a la siguiente capa.
2. Capas ocultas: estas capas realizan cálculos en los datos de entrada para aprender representaciones de los patrones y características subyacentes en los datos. Puede haber muchas capas ocultas en una red neuronal profunda.
3. Capa de salida: esta capa produce la salida final de la red, que podría ser una clasificación, predicción o algún otro tipo de decisión basada en los datos de entrada.

Se pueden utilizar diferentes tipos de capas dependiendo del problema específico que se esté resolviendo. Por ejemplo, las capas convolucionales se usan comúnmente para tareas de procesamiento de imágenes, mientras que las capas recurrentes se usan para tareas de modelado de secuencias, como el procesamiento de lenguaje natural.

El diseño y la configuración de las capas en una red neuronal profunda pueden tener un impacto significativo en el rendimiento y la capacidad de la red para aprender representaciones complejas de los datos de entrada. El proceso de selección y ajuste de las capas es una parte importante de la construcción de un modelo de aprendizaje profundo efectivo.

### 3.2.3. Arquitectura

En el contexto del DL, la arquitectura de una red neuronal se refiere a la estructura y organización de las neuronas y capas que componen el modelo. La arquitectura determina cómo las neuronas se conectan entre sí y cómo fluye la información a través de la red para realizar una tarea específica, como clasificación de imágenes o procesamiento del lenguaje natural.

Existen diversas arquitecturas de redes neuronales en el DL que han demostrado ser efectivas para diferentes tipos de problemas. Algunas de las arquitecturas más influyentes incluyen:

1. Redes Neuronales Convolucionales (CNN): Según Krizhevsky y cols. (2012) estas arquitecturas están diseñadas para el procesamiento de datos con estructura espacial, como imágenes. Las capas convolucionales permiten detectar características locales en la imagen y las capas de pooling reducen la dimensionalidad. Las CNN han sido ampliamente utilizadas en tareas de reconocimiento de imágenes y visión por computadora.
2. Redes Neuronales Recurrentes (RNN): Abordadas por Hochreiter y Schmidhuber (1997) estas arquitecturas son adecuadas para el procesamiento de datos secuenciales, como texto o series de tiempo. Las RNN tienen conexiones recurrentes que les permiten mantener una memoria a corto plazo y capturar dependencias a largo plazo en la secuencia de datos. Las variantes populares incluyen las Long Short-Term Memory (LSTM) y las Gated Recurrent Unit (GRU).
3. Redes Neuronales Tranformer: Vaswani y cols. (2017) señalan que esta arquitectura se ha vuelto prominente en el procesamiento del lenguaje natural. En lugar de utilizar conexiones recurrentes, las redes neuronales Transformer utilizan mecanismos de atención para capturar las relaciones entre las palabras en una oración o un documento. Las redes Transformer han demostrado excelentes resultados en tareas de traducción automática y generación de texto.

### 3.3. Aprendizaje superficial y profundo

El Aprendizaje Superficial o "Shallow Learnign" lo señala Schmidhuber (2015) y se refiere a técnicas de aprendizaje automático que utilizan modelos con una cantidad limitada de capas y parámetros, en contraste con el DL que se basa en redes neuronales profundas con múltiples capas y una mayor capacidad de representación de características complejas. El Shallow Learning se centra en algoritmos como las máquinas de vectores de soporte (SVM, por sus siglas en inglés) y los clasificadores lineales, los cuales son más simples y tienen un menor nivel de complejidad en

comparación con las redes neuronales profundas utilizadas en el DL.

Aunque el Shallow learning ha sido ampliamente utilizado en problemas de aprendizaje automático y ha demostrado buenos resultados en muchas aplicaciones, el DL ha ganado popularidad en los últimos años debido a su capacidad para aprender representaciones jerárquicas de los datos, lo que ha llevado a mejoras significativas en áreas como el procesamiento de imágenes, el reconocimiento de voz y la traducción automática.

Finalmente basándonos en Bengio (2009) y Schmidhuber (2015) el DL puede considerarse como una extensión del Shallow Learning al permitir que los modelos aprendan de manera automática representaciones más complejas y abstractas de los datos a través de múltiples capas de neuronas. Esto ha llevado a avances en el rendimiento y la precisión en tareas de aprendizaje automático, pero también requiere grandes conjuntos de datos de entrenamiento y un poder computacional significativo para entrenar y ajustar estos modelos más complejos

## 3.4. Componentes importantes

### 3.4.1. Capas lineales

Las capas lineales en el contexto del DL se refieren a capas de una red neuronal artificial donde las transformaciones realizadas son lineales. Estas capas se componen de neuronas que aplican una función lineal a las entradas recibidas. En otras palabras, cada neurona realiza una combinación lineal de las entradas ponderadas por los pesos correspondientes y luego aplica una función de activación no lineal para producir la salida.

En las capas lineales, la transformación realizada es una operación lineal, que se puede expresar como una multiplicación de matrices seguida de una adición de un sesgo (bias). Esta operación se puede representar matemáticamente como:

$$y = Wx + b \tag{3.1}$$

donde  $x$  es el vector de entrada,  $W$  es una matriz de pesos,  $b$  es un vector de sesgos y  $y$  es el vector de salida.

McCulloch y Pitts (1943) señala que las capas lineales son responsables de aprender los pesos y sesgos que mejor se ajustan a los datos de entrenamiento durante el proceso de entrenamiento. El número de neuronas en una capa lineal determina la dimensionalidad de la salida, y el número de capas lineales en una red contribuye a su profundidad y capacidad para aprender representaciones jerárquicas.

### 3.4.2. Funciones de activación

Las funciones de activación en las redes neuronales son funciones matemáticas que se aplican a la salida de cada neurona en una capa para introducir no linealidad en el modelo. Estas funciones son fundamentales para permitir que la red neuronal aprenda relaciones y patrones complejos en los datos.

Una función de activación comúnmente utilizada es la función ReLU (Rectified Linear Unit), definida como

$$f(x) = \max(0, x) \quad (3.2)$$

donde  $x$  es la entrada a la función. La función ReLU activa la neurona solo si su entrada es mayor que cero, de lo contrario, la salida es cero.

Otra función de activación popular es la función sigmodal, dada por

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.3)$$

Esta función comprime los valores de entrada en un rango entre 0 y 1, lo que la hace útil para problemas de clasificación binaria.



3.4.3. Funciones de pérdida

3.4.4. Capas convolucionales

3.4.5. Capas de pooling

3.4.6. Capas Recurrentes

3.4.7. Normalización por lotes (Batch Normalization)

3.4.8. Dropout

3.4.9. Entrenamiento y métodos de optimización

3.4.10. Métricas

3.4.11. Conjuntos de entrenamiento

3.4.12. Preprocesamiento de datos

3.5. Transferencia de aprendizaje (Transfer Learning)  
y ajuste fino (Fine Tuning)

## Capítulo 4

# Algoritmos de Deep Learning

### 4.1. Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales o CNN por sus siglas en inglés, es un tipo de red neuronal profunda comúnmente utilizada para tareas de reconocimiento de imágenes y videos en el aprendizaje profundo. Las CNN están diseñadas para aprender de forma automática y adaptativa jerarquías espaciales de características a partir de datos de entrada, como píxeles en una imagen.

La idea clave detrás de las CNN es usar un conjunto de filtros o núcleos que se pueden aprender para convolucionar sobre la imagen de entrada o los cuadros de video, lo que permite que la red aprenda características locales como bordes, texturas y formas. Estas características locales luego se combinan a través de capas de agrupación para crear características más abstractas y de mayor nivel que la red aprende gradualmente.

Además de las capas convolucionales y de agrupación, las CNN suelen incluir funciones de activación como ReLU (Unidad lineal rectificada) y softmax, así como capas totalmente conectadas para tareas de clasificación o regresión. La combinación de estas capas permite que la red aprenda relaciones no lineales complejas entre los datos de entrada y las predicciones de salida.

El éxito de las CNN en las tareas de reconocimiento de imágenes y videos ha llevado a su uso generalizado en muchas otras áreas de aplicación, como el procesamiento

del lenguaje natural, el reconocimiento de voz e incluso los juegos. Las CNN son una herramienta poderosa y versátil en el aprendizaje profundo que puede aprender de grandes cantidades de datos y generalizar bien a entradas nuevas e invisibles.

El aprendizaje profundo es un subconjunto del aprendizaje automático que utiliza redes neuronales artificiales con múltiples capas para modelar y resolver problemas complejos. Está inspirado en la estructura y función del cerebro humano, que es capaz de procesar grandes cantidades de información y reconocer patrones en los datos.

Los algoritmos de aprendizaje profundo pueden aprender de grandes cantidades de datos y hacer predicciones o decisiones basadas en esos datos. Estos algoritmos utilizan capas de nodos interconectados, o neuronas, para procesar y transformar los datos de entrada en representaciones cada vez más abstractas. La salida de una capa de neuronas se convierte en la entrada para la siguiente capa, lo que permite que la red aprenda características y patrones cada vez más complejos.

Uno de los beneficios clave del aprendizaje profundo es que puede descubrir automáticamente características y patrones en los datos que pueden ser difíciles o imposibles de reconocer para los humanos. Esto lo hace ideal para tareas como reconocimiento de imagen y voz, procesamiento de lenguaje natural y toma de decisiones autónoma.

Algunos de los marcos y bibliotecas de aprendizaje profundo más populares incluyen TensorFlow, PyTorch y Keras, que brindan herramientas poderosas para construir, entrenar e implementar redes neuronales profundas.

## 4.2. Data Augmentation

El aumento de datos es una técnica comúnmente utilizada en el aprendizaje profundo para aumentar artificialmente el tamaño de un conjunto de datos de entrenamiento mediante la creación de datos de entrenamiento adicionales a partir de muestras existentes. Esta técnica consiste en aplicar un conjunto de transformaciones a los datos existentes para generar nuevas muestras que sean similares pero no idénticas a los datos originales.

El objetivo del aumento de datos es mejorar la solidez y la generalización de un modelo de aprendizaje profundo exponiéndolo a datos más diversos y variados durante el entrenamiento. Esto puede ayudar al modelo a aprender a reconocer patrones y características que son relevantes en una variedad de entradas diferentes, en lugar de adaptarse demasiado a un conjunto limitado de ejemplos.

Los ejemplos de técnicas de aumento de datos incluyen:

- Recorte aleatorio o cambio de tamaño de imágenes
- Volteo horizontal o vertical de imágenes
- Rotación o corte de imágenes
- Ajustes aleatorios de brillo, contraste o saturación de imágenes
- Agregar ruido o distorsiones a las imágenes

Al aplicar este tipo de transformaciones a los datos originales, el aumento de datos puede ayudar a aumentar el tamaño efectivo de un conjunto de datos de entrenamiento y mejorar el rendimiento y la generalización de un modelo de aprendizaje profundo.

### 4.3. Redes Neuronales Recurrentes

### 4.4. Transferencia de Aprendizaje

### 4.5. TensorFlow

# Capítulo 5

## Métricas para Evaluar Modelos

Las métricas en DL son medidas utilizadas para evaluar el rendimiento de un modelo de aprendizaje automático. Las métricas proporcionan información cuantitativa sobre el rendimiento del modelo en una tarea específica, como la clasificación o la regresión. Al evaluar el rendimiento de un modelo con métricas, los investigadores y profesionales pueden comparar diferentes modelos, seleccionar el mejor modelo y optimizar los hiperparámetros del modelo para mejorar su rendimiento.

Estos son algunas de las métricas usadas para los modelos aplicados durante el desarrollo de las actividades del manual.

### 5.1. Exactitud

La exactitud (accuracy) es definida como Gu y cols. (2009):

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

Donde podemos definir:

- Verdaderos positivos (True Positive - TP): Son los valores clasificados como positivo y el valor real también es positivo.
- Falsos positivos (False Positive - FP): Son los valores clasificados como positivo y el valor real es negativo.

- Falsos negativos (False Negative - FP): Son los valores clasificados como negativo y el valor real es positivo.
- Verdaderos negativos (True Negative - TN): Son los valores clasificados como negativo y el valor real también es negativo.

La fórmula suma la función del indicador sobre todas las muestras y la divide por el número total de muestras para calcular la proporción de muestras correctamente clasificadas.

## 5.2. Precisión

Es una medida de la capacidad del modelo para predecir con precisión las muestras positivas. La métrica de precisión es definida como Gu y cols. (2009):

$$Precision = \frac{TP}{TP + FP} \quad (5.2)$$

donde TP es el número de predicciones positivas verdaderas y FP es el número de predicciones positivas falsas.

Por ejemplo, supongamos que un modelo predice la presencia o ausencia de una determinada enfermedad. Si el modelo predice que un paciente tiene la enfermedad y efectivamente la tiene, entonces es un verdadero positivo (TP). Si el modelo predice que un paciente tiene la enfermedad pero no la tiene, entonces es un falso positivo (FP). La precisión mide la proporción de verdaderos positivos entre todas las muestras que se pronosticaron como positivas, lo que puede ayudar a evaluar la capacidad del modelo para realizar predicciones positivas precisas.

Si bien la precisión es una métrica útil para evaluar el rendimiento de un modelo, debe usarse junto con otras métricas, como el recuerdo, la puntuación F1 y la sensibilidad, para obtener una comprensión más completa de las fortalezas y debilidades del modelo.

### 5.3. Sensibilidad (Recall)

Definida por Powers (2011), como el cociente entre los verdaderos positivos (TP) y la suma de los verdaderos positivos y los falsos negativos (FN):

$$Recall = \frac{TP}{TP + FN} \quad (5.3)$$

La sensibilidad proporciona información sobre la capacidad del modelo para detectar correctamente los casos positivos de una clase. Un valor alto de sensibilidad indica que el modelo es capaz de identificar la mayoría de los casos positivos, mientras que un valor bajo indica que el modelo tiende a perder algunos casos positivos.

### 5.4. Matriz de Confusión

Sokolova y Lapalme (2009) define una matriz de confusión como una herramienta utilizada para evaluar el desempeño del modelo al predecir las clases de un conjunto de datos. La matriz muestra el número de muestras que son clasificadas correctamente e incorrectamente para cada clase.

La matriz de confusión tiene una estructura de tabla con filas y columnas correspondientes a las clases reales y las clases predichas, respectivamente. A continuación se presenta la notación comúnmente utilizada para una matriz de confusión binaria:

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (5.4)$$

- Verdaderos positivos (True Positive - TP): Son los valores clasificados como positivo y el valor real también es positivo.
- Falsos positivos (False Positive - FP): Son los valores clasificados como positivo y el valor real es negativo.
- Falsos negativos (False Negative - FN): Son los valores clasificados como negativo y el valor real es positivo.

- Verdaderos negativos (True Negative - TN): Son los valores clasificados como negativo y el valor real también es negativo.

La matriz de confusión proporciona una visión general del rendimiento del modelo en términos de aciertos y errores de clasificación para cada clase.



# Capítulo 6

## Metodología de las Prácticas

Durante la impartición del curso se diseñaron dos tipos de prácticas; Practicas de Laboratorio impartidas en clase y Actividades Complementarias para evaluar el progreso y entendimiento de los alumnos con respecto a los temas abarcados. Todas las prácticas presentadas en el anexo contienen las siguientes fases, en el orden en que se muestran:

1. Objetivo de la práctica.
2. Conceptos.
3. Herramientas a usar.
4. Desarrollo.
  - a)* Entender el Problema.
  - b)* Definir un criterio de evaluación.
  - c)* Preparar los datos.
  - d)* Construir el modelo.
  - e)* Análisis de errores.
  - f)* Implementación.

En el caso especial de la práctica 5 de laboratorio (Uso de TensorFlow Distribuido) no se aplican los puntos .... (tiene caso mencionar esto si se realizó un ejemplo?)

N° Práctica	NOMBRE	DATASET	CRITERIO DE EVALUACIÓN	DESCRIPCIÓN
1	Clasificación Binaria y Redes Convolucionales Básico	Mosquitos Aedes y Aegyptys	Exactitud	Crear una red neuronal básica para clasificar dos especies diferentes de mosquitos analizando las imágenes proporcionadas.
2	Clasificación Binaria y Redes Convolucionales Avanzado	COVID-19 Xray Dataset (Train & Test Sets)	Exactitud	Aplicar lo aprendido en la actividad de los mosquitos pero aplicado a imágenes de Rayos X de diferentes pacientes.
3	Redes Convolucionales y Data Augmentation	100 Sports Image Classification	Precisión y Sensibilidad.	Aprender a expandir el dataset cuando el conjunto de datos es limitado y crear un modelo de convolución multiclase para todos los deportes.
4	Redes Neuronales Recurrentes	News Headlines Dataset For Sarcasm Detection	Exactitud y Precisión	Hacer análisis de texto, seleccionar las columnas más relevantes y crear un modelo capaz de identificar patrones y contextos.
5	Uso de TensorFlow Distribuido	3 nodos virtuales con CPU y GPU c/u	No aplica	Mostrar cómo se realiza la instalación de TensorFlow y cómo se comunican los nodos para el procesamiento paralelo de grandes volúmenes de datos.
6	Transferencia de aprendizaje	Game of Deep Learning: Ship datasets	Precisión y matriz de confusión	Aprenderán a usar redes preentrenadas para aplicar entrenamientos más pesados a modelos con clasificaciones más complejas

Tabla 6.1: Tabla de Prácticas de Laboratorio con su contenido

N° Práctica	NOMBRE	DATASET	CRITERIO DE EVALUACIÓN	DESCRIPCIÓN
1	Clasificación Multiclase de Letras	A-Z Handwritten Alphabets in .csv format	Exactitud	El alumno debe entrenar un modelo a partir del dataset proporcionado y comprobar su efectividad haciendo predicción de una frase escrita a mano.
2	Análisis de Sentimientos de Tweets	Twitter US Airline Sentiment	Exactitud	El alumno debe hacer un análisis del dataset para decidir qué información es necesaria de procesar, desarrollará un modelo a partir de ello y hará una prueba con un texto de su autoría.
3	Sarcasmo en Noticias Distribuido	News Headlines Dataset For Sarcasm Detection	Exactitud y Precisión, Velocidad de ejecución	El alumno debe probar el TensorFlow ejecutando la práctica vista en clase en uno de los nodos del servidor y ejecutarlo de manera paralela en los tres nodos del mismo.
4	Creación de Red Neuronal	Elección del alumno	Elección del alumno	El alumno debe mostrar lo aprendido a lo largo del curso creando su propia red neuronal para realizar un análisis completo de un dataset de su elección y mostrar resultados concretos.

Tabla 6.2: Tabla de Actividades Complementarias con su contenido

## Capítulo 7

### Resultados y Discusión

\*\*\**PENDIENTE*\*\*\*

# Apéndice

Los anexos consisten en:

- Libretas de JupyterLab programadas en Python de las prácticas desarrolladas.
- Enlaces y referencias de los datos originales usados para las prácticas desarrolladas.

## .1. Prácticas

Las prácticas, enlaces a grabaciones y libretas de JupyterLab se encuentran en el siguiente repositorio en la plataforma

**GitHub:** TBD

## .2. Datos

Los conjuntos de datos que se usaron en las prácticas mostradas anteriormente, se pueden descargar de los enlaces siguientes:

- Mosquitos Aedes y Aegyptys.  
<http://basurae.iies.unam.mx/webmosquito/html/>
- COVID-19 Xray Dataset (Train & Test Sets)  
<https://www.kaggle.com/khoongweihao/covid19-xray-dataset-train-test-sets>
- 100 Sports Image Classification  
<https://www.kaggle.com/gpiosenka/sports-classification>
- Conjunto de Datos de Encabezados de Noticias para la Detección de Sarcasmo  
<https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection/code>
- Game of Deep Learning: Ship datasets  
<https://www.kaggle.com/datasets/arpitjain007/game-of-deep-learning-ship-datasets>
- A-Z Handwritten Alphabets in .csv format  
<https://www.kaggle.com/datasets/sachinpatel21/az-handwritten-alphabets-in-csv-format>

- Twitter US Airline Sentiment

<https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment>

# Referencias

- Alpaydin, E. (2010). *Introduction to machine learning*. MIT Press.
- Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and Trends® in Machine Learning*, 2(1), 1–127.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., y Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)
- Gu, Q., Zhu, L., y Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. *Communications in Computer and Information Science*, 461-471.
- Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hochreiter, S., y Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Krizhevsky, A., Sutskever, I., y Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. En *Advances in neural information processing systems* (pp. 1097–1105).
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

- McCulloch, W. S., y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Ng, A. (2017). *Machine learning yearning*. Gradient descent.
- Powers, D. M. W. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.
- Serre, T. (2018). Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science*, 4, 381–405.
- Shalev-Shwartz, S., y Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Sokolova, M., y Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. En *Advances in neural information processing systems* (pp. 5998–6008).