



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Inteligencia Artificial: Un
enfoque práctico desde la
experiencia de Q&A Data
Scientist**

INFORME DE ACTIVIDADES PROFESIONALES

Que para obtener el título de

Ingeniero Industrial

P R E S E N T A

Alejandro Herrera Botello

ASESOR DE INFORME

M.C. Raúl Gilberto Valdez Navarro



Ciudad Universitaria, Cd. Mx., 2024

Inteligencia Artificial: Un enfoque práctico desde la experiencia en Q&A Data Scientist

Índice

I. INTRODUCCIÓN	4
II. PLANTEAMIENTO DEL PROBLEMA	5
III. OBJETIVOS	6
OBJETIVO GENERAL:.....	6
OBJETIVOS ESPECÍFICOS:.....	6
1. MARCO DE REFERENCIA	6
1.1 ¿Quién o qué es Shift Technology?.....	6
1.2 Política de Shift Technology.....	7
1.3 Misión.....	7
1.4 Visión.....	7
1.5 Valores.....	7
1.6 Principios.....	8
1.7 ¿Qué es la Inteligencia Artificial?.....	8
1.8 ¿Qué es el Aprendizaje Automático?.....	8
1.9 Organigrama.....	9
1.1.0 Mis actividades laborales como Q&A Data Scientist.....	9
1.1.2 Principales actividades realizadas:.....	10
2. DEFINICIÓN DEL PROBLEMA	11
3. EL ABC DEL MACHINE LEARNING	13
3. ANÁLISIS DEL PROBLEMA Y SOLUCIÓN IMPLEMENTADA	28
3.1 Análisis del problema.....	30
3.2 Solución implementada.....	30
3.3 Metodología.....	32
4. RESULTADOS	35
5. CONCLUSIONES	37
BIBLIOGRAFÍA Y REFERENCIAS	38

I. INTRODUCCIÓN

Durante mi tiempo como Q&A Data Science en SHIFT TECHNOLOGY, una empresa líder a nivel internacional en Inteligencia Artificial, reconocida por su innovación en la detección de fraudes y la automatización de reclamos, tuve la oportunidad de aprender y aplicar una variedad de conocimientos y habilidades en el ámbito profesional. SHIFT TECHNOLOGY se destaca por su desarrollo continuo de algoritmos avanzados que permiten a las compañías de seguros mejorar significativamente la calidad y eficiencia de sus procesos. Esta experiencia me permitió desarrollar una comprensión profunda de cómo la Inteligencia Artificial y el aprendizaje automático pueden ser aplicados en la industria de seguros para optimizar la detección de fraudes y agilizar la gestión de reclamos.

Uno de los aspectos más interesantes de mi trabajo en SHIFT TECHNOLOGY fue la oportunidad de trabajar en proyectos desafiantes y emocionantes. En particular, tuve la oportunidad de colaborar en el desarrollo de modelos de detección de fraude basados en técnicas de aprendizaje automático. Estos modelos se utilizaron para identificar patrones de comportamiento sospechosos en las reclamaciones de seguros o para automatizar procesos digitales empresariales, lo que permitió a las aseguradoras tomar medidas preventivas y reducir los costos asociados con el fraude.

Para desarrollar estos modelos, tuve que aplicar una variedad de técnicas de aprendizaje automático, desde la selección y limpieza de datos hasta el entrenamiento y validación de modelos. También tuve que trabajar en estrecha colaboración con otros especialistas en datos y miembros del equipo de desarrollo de productos, para asegurarme de que los modelos fueran precisos y fáciles de usar para los clientes.

Además de mis habilidades técnicas, también aprendí a comunicarme de manera efectiva con otros miembros del equipo en un ámbito internacional en 3 diferentes idiomas (inglés, francés y español), incluyendo a los gerentes de producto y a los especialistas en datos de diferentes áreas de la empresa. Esto me permitió entender las necesidades de los clientes y desarrollar modelos que satisficieran sus necesidades.

En este escrito, compartiré algunos de los aprendizajes clave que adquiriré durante mi tiempo en SHIFT TECHNOLOGY, así como las lecciones aprendidas sobre cómo aplicar la tecnología de manera efectiva para resolver problemas empresariales complejos en la industria de seguros.

II. PLANTEAMIENTO DEL PROBLEMA

En la actualidad, la industria de detección de fraude representa un nicho estratégico en el que las empresas están invirtiendo significativas sumas de capital. Para abordar este desafío, diversas empresas han establecido organismos internos dedicados a la detección de fraudes. Sin embargo, Shift Technology se destaca en este campo gracias a su enfoque innovador y al desarrollo continuo de tecnologías avanzadas.

La empresa ha logrado posicionarse como líder a nivel internacional en la detección de fraudes y en la automatización de reclamos, gracias a la implementación de algoritmos cada vez más precisos y eficientes. Este éxito se debe en gran medida a su capacidad para adaptarse rápidamente a las nuevas tendencias del mercado y para desarrollar soluciones innovadoras que satisfacen las demandas cambiantes de sus clientes.

A pesar del reconocimiento y la reputación de Shift Technology en el ámbito de la detección de fraudes, el sector sigue enfrentando diversos desafíos. Entre estos desafíos se incluyen la evolución constante de las tácticas de fraude, la complejidad creciente de los patrones fraudulentos y la necesidad de mantener la precisión y eficiencia de los algoritmos de detección en un entorno en constante cambio.

En este contexto, surge la necesidad de analizar en profundidad las estrategias y tecnologías empleadas por Shift Technology, así como de identificar áreas de mejora y oportunidades para fortalecer aún más su posición como líder en el campo de la detección de fraudes y la automatización de reclamos.

III. OBJETIVOS

OBJETIVO GENERAL:

- Describir las principales funciones que desarrolla un Q&A Data Scientist en la industria de detección de fraude y Automation Claim con herramientas tales como la Inteligencia Artificial.

OBJETIVOS ESPECÍFICOS:

- Describir las principales tareas en proyectos de detección de fraude
- Describir cómo funciona la industria de detección de fraude en el mundo.
- Presentar resultados obtenidos en diferentes proyectos de Automation Claim y detección de fraude.
- Definir áreas de oportunidad para mejorar la productividad.

1. MARCO DE REFERENCIA

1.1 ¿Quién o qué es Shif Thechnology?

Shift Technology es una empresa de tecnología con sede en París, Francia, que se especializa en el desarrollo de soluciones de Inteligencia Artificial y aprendizaje automático para la industria de seguros. La empresa fue fundada en 2013 por un equipo de expertos en análisis de datos, fraude y tecnología, y desde entonces ha crecido rápidamente hasta convertirse en un líder mundial en la detección de fraudes en seguros.

Shift Technology utiliza técnicas avanzadas de análisis de datos, aprendizaje automático y procesamiento del lenguaje natural para ayudar a las compañías de seguros a identificar y prevenir el fraude de seguros. Sus soluciones se basan en modelos de aprendizaje automático para analizar grandes cantidades de datos y detectar patrones sospechosos en las

reclamaciones de seguros además de brindar productos innovadores como digitalización y automatización de procesos.

La empresa ha recibido una importante inversión de capital de riesgo, y ha recaudado más de 100 millones de dólares de inversores de renombre como Accel, General Catalyst e Iris Capital, entre otros. Shift Technology ha sido reconocida por varias organizaciones y premios, incluyendo el premio a la Mejor Startup en la categoría de Inteligencia Artificial y Aprendizaje Automático en los premios de Fintech France 2020. Además, Shift Technology se enorgullece de su cultura de trabajo colaborativo, diverso e inclusivo, y de su compromiso con la innovación continua y la excelencia técnica. Shift Technology es una empresa líder en el campo de la tecnología de detección de fraude en seguros, con un enfoque en la colaboración, la personalización y la innovación.¹

1.2 Política de Shift Technology

Liberar a las aseguradoras para estar más presentes para más clientes en más formas que nunca.

1.3 Misión

Imaginar e inventar soluciones que transformen las decisiones en seguros para el beneficio de todos.

1.4 Visión

Una industria aseguradora en la que las mejores decisiones conduzcan a oportunidades y resultados sorprendentes.

1.5 Valores

Unión, fuerza, inclusión, multiculturalidad, respeto, innovadores, efectivos y eficientes.

Shift Technology, Series D Investment Round Led by Advent International, 2021, <https://www.shift-technology.com/resources/news/shift-technology-secures-series-d#:~:text=PARIS%20and%20BOSTON%20%E2%80%93%20May%202006,investment%20round%20of%20%24220%20million.>

1.6 Principios

La empresa se esfuerza por ser un socio confiable y estratégico para las compañías de seguros, y para ayudar a transformar la industria de seguros en una industria más eficiente, transparente y centrada en el cliente. Shift Technology se enfoca en la innovación continua y la mejora constante de sus soluciones para ofrecer el máximo valor a sus clientes.

1.7 ¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial (IA) es una rama de la informática que busca desarrollar sistemas capaces de realizar tareas cognitivas complejas, como la percepción, el razonamiento y el aprendizaje autónomo." - UNAM, Centro de Investigación en Computación.

"La inteligencia artificial (IA) es un campo de investigación en la informática que se enfoca en la creación de máquinas capaces de realizar tareas que requieren inteligencia humana, como el aprendizaje, la toma de decisiones y la comprensión del lenguaje natural. La IA se basa en algoritmos de aprendizaje automático que permiten a las máquinas aprender de los datos y mejorar su desempeño con el tiempo." - Alpaydin, Introducción al aprendizaje automático.

1.8 ¿Qué es el Aprendizaje Automático?

Según Alpaydin, el aprendizaje automático es "un conjunto de técnicas que permiten a los sistemas computacionales mejorar automáticamente a través de la experiencia" (p. 1). Alpaydin, E. (2010). Introduction to machine learning. Cambridge, MA: MIT Press.

1.9 Organigrama

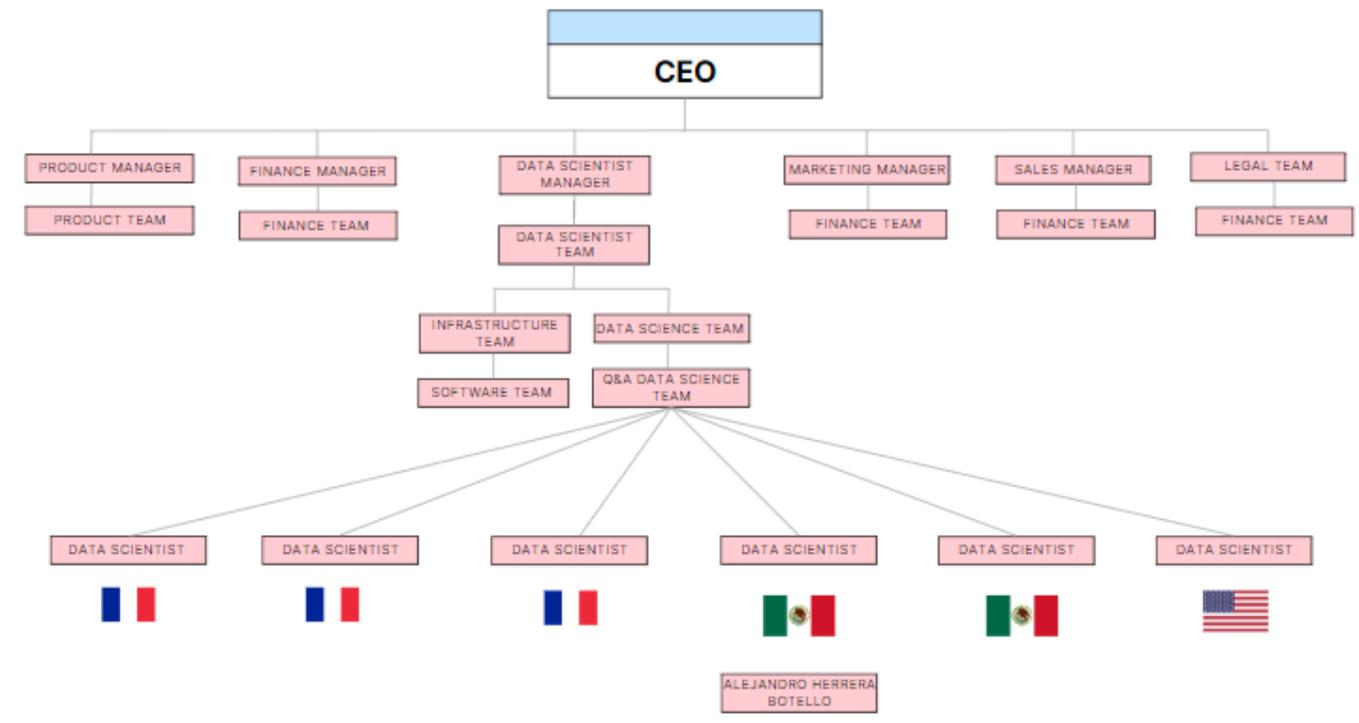


Figura 1. Organigrama

1.1.0 Mis actividades laborales como Q&A Data Scientist

El trabajo de un Q&A Data Scientist implica ayudar a las empresas a aprovechar al máximo sus datos mediante el análisis y el aseguramiento de la calidad de los datos y metas en específico. En este rol, se debe tener una estrecha colaboración con diferentes equipos de la empresa, incluyendo marketing, ventas, y operaciones, para recopilar, procesar y analizar datos que ayuden a resolver problemas comerciales específicos.

El Q&A Data Scientist también tiene la responsabilidad de crear modelos de datos y algoritmos que puedan responder preguntas de la empresa en tiempo real, lo que requiere habilidades avanzadas en programación y estadística. Además, este rol implica la capacidad de comunicar

los resultados de los análisis de datos de manera clara y concisa a diferentes audiencias dentro de la empresa, incluyendo a los equipos ejecutivos y no técnicos.

En general, el trabajo de un Q&A Data Scientist es muy importante en el mundo empresarial actual, ya que las empresas están cada vez más interesadas en aprovechar al máximo sus datos para tomar mejores decisiones comerciales y mejorar su competitividad. Como tal, un profesional de datos exitoso en este rol debe ser capaz de combinar habilidades técnicas avanzadas con la capacidad de comunicarse eficazmente con diferentes partes interesadas en la empresa.

1.1.2 Principales actividades realizadas:

Comprender las necesidades de negocio: Entender las necesidades y objetivos comerciales de la empresa es clave para un correcto desarrollo en cualquier tipo de industria. Esto significa trabajar estrechamente con los equipos internos para comprender las preguntas clave que la empresa necesita responder a través del análisis de datos.

Recopilar y procesar los datos: Una vez que entiende las preguntas que la empresa necesita responder, debo recopilar y procesar los datos necesarios. Esto implica la capacidad de trabajar con una amplia variedad de fuentes de datos, incluyendo datos estructurados y no estructurados, y la capacidad de limpiar y normalizar los datos para su uso en análisis.

Analizar los datos: Una vez que los datos estén procesados y listos para su análisis, debo utilizar herramientas y técnicas avanzadas para realizar el análisis de datos. Esto incluye la creación de modelos de datos y la implementación de algoritmos avanzados de aprendizaje automático para analizar los datos y obtener información significativa.

Comunicar los resultados: Una vez analizados los datos, debo comunicar los resultados de manera clara y concisa a diferentes audiencias dentro de la empresa, incluyendo a los equipos ejecutivos y no técnicos. Esto requiere habilidades avanzadas de presentación y la capacidad de traducir los resultados técnicos en un lenguaje comprensible para diferentes audiencias.

Mejorar continuamente: El análisis de datos es un proceso continuo, por lo que debo estar comprometido con la mejora continua y la optimización de los modelos y algoritmos utilizados

para el análisis. Esto implica estar al día con las últimas tendencias y técnicas en el campo de la ciencia de datos y la Inteligencia Artificial.

2. DEFINICIÓN DEL PROBLEMA

GEICO, una destacada empresa en el sector de seguros, ha reconocido la necesidad de modernizar sus procesos de reclamaciones para ofrecer a sus clientes una experiencia más eficiente y efectiva. El problema radica en que sus métodos actuales implican la necesidad de que los clientes visiten las oficinas físicas o interactúen telefónicamente de manera ineficiente para gestionar sus reclamaciones. Este proceso obsoleto no solo resulta inconveniente para los asegurados, sino que también supone un desafío para la empresa en términos de eficiencia y rentabilidad.

En la era digital actual, los asegurados esperan una experiencia de servicio más ágil y accesible. Los métodos tradicionales de presentación de reclamaciones, que requieren visitas personales o llamadas telefónicas prolongadas, no solo son inconvenientes, sino que también generan costos operativos adicionales para GEICO. Además, la necesidad de procesar una gran cantidad de documentación en papel aumenta la posibilidad de errores y retrasos en el proceso de reclamaciones.

Para abordar este desafío, GEICO ha decidido tomar medidas proactivas para modernizar su proceso de reclamaciones. Como parte de esta iniciativa, la empresa ha contratado a Shift Technology, una compañía líder en Inteligencia Artificial y aprendizaje automático aplicado a la industria de seguros.

Dentro de este proyecto, se debe asegurar que la automatización de reclamaciones se desarrolle de manera efectiva y eficiente. Esto implica colaborar estrechamente con el equipo de Shift Technology y el personal de GEICO para:

1. Identificar los flujos de trabajo actuales de reclamaciones y analizar cómo se pueden digitalizar y automatizar de manera óptima.
2. Desarrollar un sistema que permita a los asegurados presentar sus reclamaciones de manera electrónica, cargar documentos relevantes y realizar seguimiento en tiempo real del estado de sus reclamaciones.
3. Implementar medidas de seguridad para garantizar la confidencialidad y la integridad de los datos del cliente.
4. Garantizar que la nueva solución sea fácil de usar y esté alineada con las expectativas de los asegurados.

Un Paso Más Allá: Detección de Fraude:

Además de modernizar el proceso de reclamaciones, esta solución tiene un beneficio adicional. Los datos recopilados a través del nuevo sistema de automatización de reclamaciones servirán como entrada a otro algoritmo de inteligencia artificial desarrollado por Shift Technology: la detección de fraude. Esto permitirá a GEICO identificar patrones sospechosos y minimizar el riesgo de fraudes en las reclamaciones.

Mi rol principal será desarrollar un sistema de aprendizaje automático (machine learning) especializado en la correcta clasificación de documentos. Este sistema permitirá identificar y etiquetar de manera precisa los diferentes tipos de documentos, como facturas, tickets, recibos, entre otros. Mi enfoque se centrará en implementar algoritmos y técnicas de aprendizaje supervisado para que el modelo pueda aprender de un conjunto de datos etiquetado y, posteriormente, realizar clasificaciones precisas en tiempo real.

En la clasificación de documentos, existen varios tipos de algoritmos de aprendizaje automático efectivos, como las Máquinas de Soporte Vectorial (SVM), que son excelentes para separar categorías en datos de alta dimensionalidad como texto²; Convolutional Neural Networks (CNN) para procesamiento de imágenes de documentos³; el Naïve Bayes, adecuado para tareas de

² SVM Tutorial, SVM - Understanding the math, <https://www.svm-tutorial.com/>

³ LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature*, 521(7553), 436-444. DOI: 10.1038/nature14539.

clasificación de texto⁴; la Regresión Logística, simple y efectiva para problemas binarios o multiclase⁵; y Random Forest , un ensamble de árboles de decisión que es robusto para clasificar documentos debido a su capacidad para manejar múltiples características y clases⁶. Estos algoritmos son esenciales para tareas de clasificación de documentos en una variedad de campos, desde procesamiento de texto hasta análisis de imágenes.

En este proyecto de clasificación de documentos, opté por utilizar el aprendizaje supervisado debido a que la empresa cuenta con un equipo especializado para el etiquetado de datos en el que yo mismo participé. Esto significa que tenemos ejemplos claros de cada tipo de documento con sus respectivas etiquetas, lo que nos permite entrenar un modelo de machine learning para que aprenda a clasificar documentos de manera precisa.

En contraste, el aprendizaje no supervisado habría sido menos adecuado en este contexto. En el aprendizaje no supervisado, el modelo no se guía por etiquetas predefinidas y, en lugar de eso, busca patrones y similitudes en los datos. Esto puede generar agrupamientos o categorías que no necesariamente se corresponden con los tipos de documentos específicos que deseamos clasificar puesto que para cierto tipo de documento aún existe demasiada variabilidad como nombre, datos personales, datos sensibles y resultados diferentes. Además, la interpretación y asignación de significado a estos agrupamientos podrían requerir una intervención humana adicional.

3. EI ABC DEL MACHINE LEARNING

Para comprender un tema de tal complejidad, es imperativo tener un sólido conocimiento de las bases de la Inteligencia Artificial, dado que existen diversos tipos y aplicaciones dentro de este campo. La evolución exponencial en la capacidad de cálculo y memoria de las computadoras desde su creación ha impulsado a la humanidad a desarrollar algoritmos que simulan el proceso de aprendizaje humano. Este punto de convergencia da origen al Machine Learning o

⁴ Monkeylearn, Naive Bayes Classifier Explained, <https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/>

⁵ Towardsdatascience, Logistic Regression Explained, <https://towardsdatascience.com/understanding-logistic-regression-9b02c2aec102>

⁶ Towardsdatascience, Understanding Random Forests, <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

aprendizaje automático, un fenómeno que, en la actualidad, se manifiesta en la capacidad de Inteligencia Artificial para replicar voces humanas, generar imágenes realistas de individuos realizando acciones ficticias, componer canciones con una voz artificial o replica de una voz natural, llevar a cabo el reconocimiento facial, entre otras aplicaciones aparentemente infinitas.

El Machine Learning se define como el conjunto de técnicas mediante las cuales un algoritmo tiene la capacidad de aprender a partir de los datos con los que cuenta, ya sea el conjunto de datos con los que fue inicialmente creado y su desempeño o con base en un resultado deseado⁷. Esta tecnología, en constante evolución, prescinde de la programación convencional, permitiendo que el algoritmo aprende y se ajuste de manera autónoma a partir de la información proporcionada.

El aprendizaje automático ofrece diversas aproximaciones para abordar distintos tipos de tareas y desafíos. Uno de los paradigmas más comunes es el aprendizaje supervisado, donde el modelo se entrena con un conjunto de datos etiquetado, aprendiendo así a mapear entradas a salidas específicas⁸. Un ejemplo claro es la clasificación de imágenes, en donde se etiqueta un dataset y el algoritmo aprende a identificar ciertas características y patrones.

En contraste, el aprendizaje no supervisado trabaja con conjuntos de datos sin etiquetas, buscando patrones o estructuras latentes. Aquí, técnicas como el clustering permiten agrupar datos similares, como clientes con patrones de compra afines. Además, la reducción de dimensionalidad es un enfoque que conserva información esencial al disminuir la cantidad de características⁹. Otro enfoque es el aprendizaje por refuerzo, donde un agente interactúa con un entorno y toma decisiones para maximizar recompensas a largo plazo. Este método se aplica en juegos, robótica y otras áreas donde la toma de decisiones secuenciales es clave¹⁰.

⁷ Divason, J., Martinez De Pison, F.J., Romero, A., & Saenz-De-Cabezón, E. (2023). "Artificial Intelligence Models for Assessing the Evaluation Process of Complex Student Projects." IEEE Transactions on Learning Technologies.

⁸ Mitchell, T. M. (1997). "Machine Learning." McGraw Hill, New York.

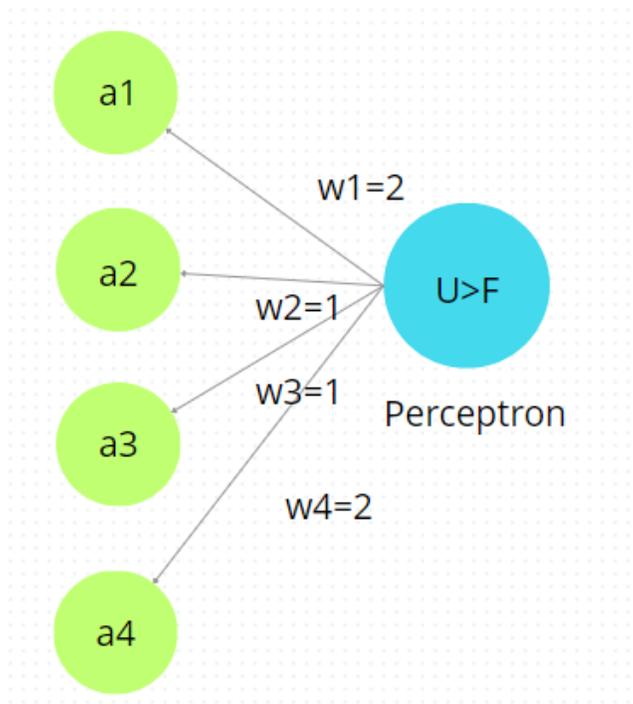
⁹ Bishop, C. M. (2006). "Pattern Recognition and Machine Learning." Springer.

¹⁰ Sutton, R. S., & Barto, A. G. (2018). "Reinforcement Learning: An Introduction." The MIT Press.

Adicionalmente, el aprendizaje semisupervisado combina ejemplos etiquetados y no etiquetados para mejorar la eficiencia del modelo. Este enfoque es útil cuando la obtención de datos etiquetados es costosa o limitada¹¹.

Estos enfoques no son mutuamente excluyentes; la elección depende de la naturaleza del problema y la disponibilidad de datos etiquetados. La versatilidad del aprendizaje automático se manifiesta en su capacidad para abordar diferentes problemas, desde la clasificación de imágenes hasta la optimización de decisiones en entornos complejos.

Para aplicar estos tipos de aprendizajes existen muchas técnicas(algoritmos), pero en las que mi proyecto se enfocará es en el uso de redes neuronales, una de las más utilizadas en machine learning. Las redes neuronales son modelos computacionales inspirados en la estructura y funcionamiento del cerebro humano. En una red neuronal, el perceptrón es la unidad básica. Cada perceptrón tiene entradas ($a_1, a_2, a_3 \dots \dots \dots a_n$) que representan características o variables, y asigna pesos ($w_1, w_2, w_3 \dots \dots \dots w_n$) a estas entradas, que básicamente son las conexiones entre diferentes elementos como se muestra en la Figura 2.



¹¹ Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). "Semi-Supervised Learning." The MIT Press.

Figura 2. Red neuronal 1

Imagen inspirada de los artículos de 3blue1brown

Referencia: (3blue1brown, 2017)

<https://www.3blue1brown.com/topics/neural-networks>

Además, hay un valor de umbral (t) y una función de activación que determina si el perceptrón se activa o no. La función de activación introduce no linealidades en el modelo, permitiendo a la red aprender patrones más complejos. Las funciones de activación serán tema que explicaré más adelante por ahora sigamos el siguiente ejemplo:

Cuando nos enfrentamos a la decisión de adquirir un nuevo automóvil, es posible modelar este proceso mediante una red neuronal con cuatro variables de entrada: a_1 =disponibilidad de capital, a_2 =disponibilidad del modelo deseado, a_3 =disponibilidad del color deseado, y a_4 =tiempo de entrega deseado. En este escenario, nuestro valor de umbral es establecido en 5, y cada variable de entrada a_n solo puede tener un valor de 0 o 1, indicando la presencia o ausencia de disponibilidad.

Supongamos que asignamos pesos específicos a estas variables: $w_1=2$, $w_2=2$, $w_3=1$, $w_4=1$. Esto significa que la disponibilidad de capital y el modelo deseado tienen un peso mayor en la decisión de compra. Si, por ejemplo, se activa la disponibilidad de capital y otras entradas, la suma ponderada sería $2(1)$ disponibilidad de capital + $2(1)$ disponibilidad del modelo + $1(1)$ disponibilidad del color + $1(1)$ tiempo de entrega. En este escenario hipotético, si esta suma supera el valor de umbral de 5, nuestra función de activación se activaría, simulando así una compra exitosa de un automóvil, como se muestra en la Figura 3.

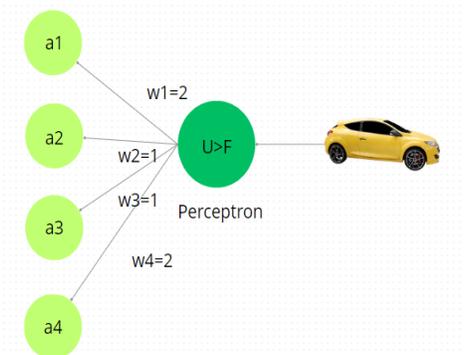


Figura 3. Red neuronal 2

Imagen inspirada de los artículos de 3blue1brown

Referencia: (3blue1brown, 2017)

<https://www.3blue1brown.com/topics/neural-networks>

Este proceso adquiere aún más fascinación al introducir pesos a nuestras variables de entrada. Al asignar diferentes pesos y la posibilidad de activar perceptrones adicionales, generando capas de abstracción y complejidad en la red neuronal. Podemos imaginar que la activación de un perceptrón se convierte en la entrada para otro perceptrón, y así sucesivamente. Esta cascada de activaciones permite simular problemas complejos con miles de entradas, donde ciertos patrones activan otros patrones, generando una representación más interesante y adaptativa del proceso de toma de decisiones, como se muestra en la Figura 4.

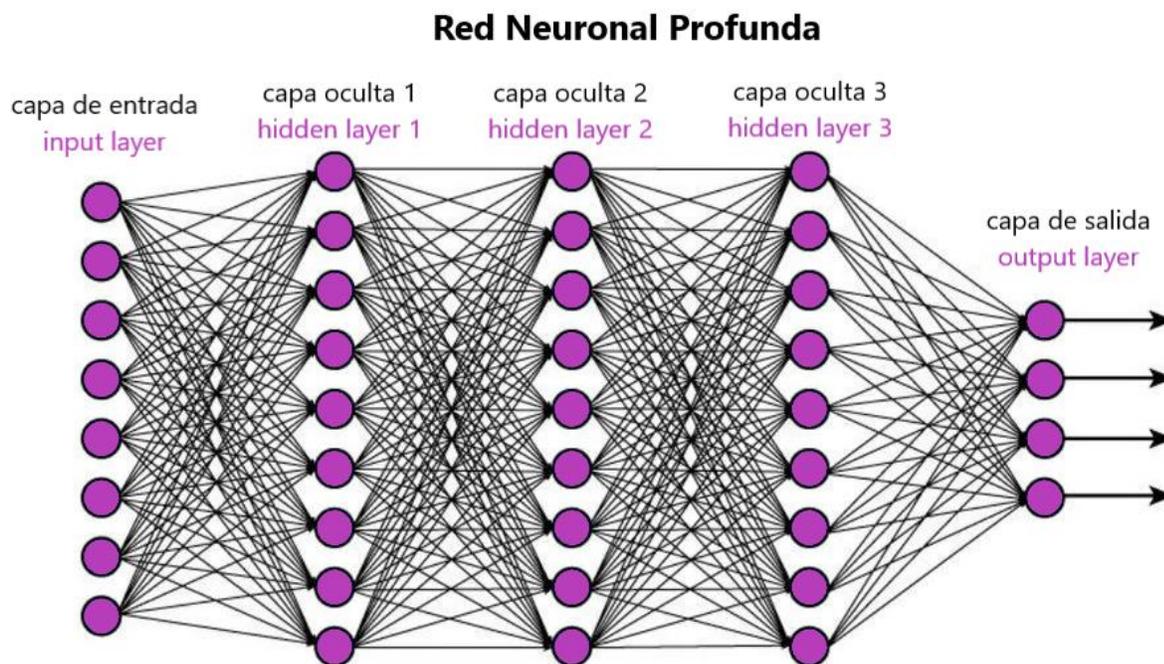


Figura 4. Arquitectura de una red con múltiplo capas

Referencia: (sitiobigdata.com, 2020)

<https://sitiobigdata.com/2019/12/24/redes-neuronales-profundas/>

En la red neuronal de la Figura 4, se aprecian tres subcapas, donde los patrones de activación impulsan la activación de la siguiente subcapa de manera sucesiva. La elección del número de subcapas y la cantidad de neuronas o perceptrones necesarios en la red dependen del problema específico a abordar. En la actualidad, en la industria de la inteligencia artificial, ya existen redes preestablecidas. En un entorno corporativo, la práctica común implica la adaptación de estas redes específicas al problema actual, evitando así iniciar el proceso desde

cero. Este enfoque permite aprovechar la experiencia y la eficacia de arquitecturas de red previamente desarrolladas, agilizando el desarrollo y la implementación de soluciones en el ámbito de la Inteligencia Artificial.

Ahora, supongamos que deseamos desarrollar una red neuronal capaz de aprender a identificar imágenes, un desafío indudablemente complejo en el que nuestro cerebro humano es un experto natural en el reconocimiento de patrones visuales. Si aspiramos a emular esta función mediante una red neuronal, ¿cómo podríamos empezar? Un enfoque inicial efectivo sería asignar a nuestra primera subcapa la tarea de reconocer líneas, bordes o semicírculos en una imagen. La segunda subcapa, por su parte, podría especializarse en identificar círculos, interpretando estos como la combinación de semicírculos previamente detectados en la subcapa anterior. En este contexto, cada capa de la red dependería de la abstracción específica que buscamos obtener del mundo real, permitiendo así que la red aprenda a reconocer patrones visuales cada vez más complejos, como se observa en la Figura 5.

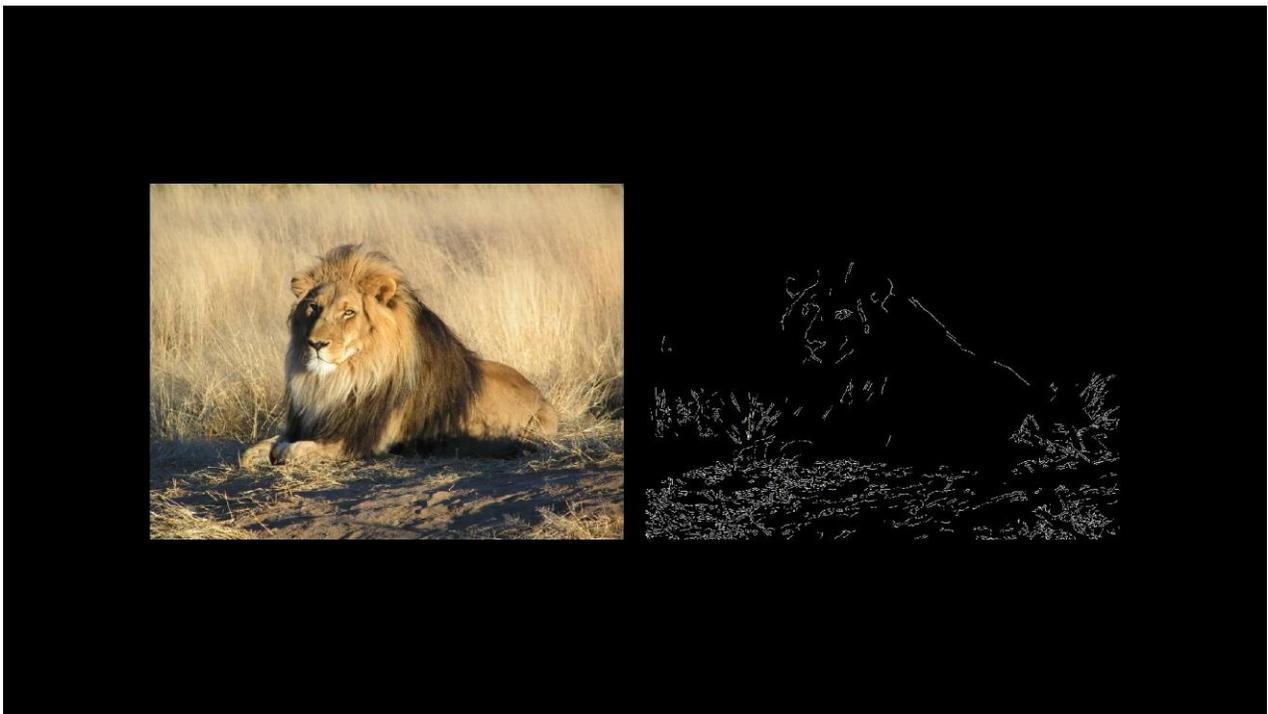


Figura 5. Reconocimiento de bordes en una imagen

Referencia: (3blue1brown, 2017)

<https://www.3blue1brown.com/lessons/neural-networks>

Ahora bien, si en nuestra imagen deseamos identificar ciertos bordes o patrones, asignaríamos pesos en aquellas áreas donde consideramos que dichos bordes podrían aparecer, otorgándoles valores negativos alrededor de la región de interés. De esta manera, la suma ponderada, como se muestra en la ecuación 1, se optimizará cuando el modelo detecte el patrón deseado en la imagen, facilitando así el reconocimiento¹².

$$a_1w_1 + a_2w_2 + a_3w_3 + a_4w_4 + \dots \dots \dots a_nw_n \quad EC. (1)$$

La suma ponderada en nuestra red neuronal puede adoptar cualquier valor, por lo cual se agrega un término constante, un número real, para tener un control sobre dicha suma y activación, como se muestra en la ecuación 2.

$$a_1w_1 + a_2w_2 + a_3w_3 + a_4w_4 + \dots \dots \dots a_nw_n + \textit{Sesgo} \quad EC. (2)$$

Aquí surge un dilema: aunque esta práctica resulta en un comportamiento predecible y real, independientemente del número de entradas o la complejidad de la red neuronal, esta simplicidad conlleva una limitación. Una red con un comportamiento lineal, lamentablemente, no puede abordar problemas no lineales ni simular funciones complejas.

Para superar esta limitación, compactamos toda nuestra suma en una función. En la industria actual, existen diversos tipos de funciones de activación, y la elección de una en particular depende del problema que se está resolviendo. Sin embargo, el objetivo es unánime: activar nuestra neurona de manera no lineal. Un ejemplo ilustrativo es la función de activación sigmoide, como se observa en la Figura 6, donde valores muy positivos conducen a una salida de 1, mientras que valores muy negativos resultan en una salida de 0. Este enfoque no lineal es esencial para que la red pueda abordar de manera efectiva problemas complejos y no lineales en su funcionamiento.

¹² "3Blue1Brown." Neural Networks, Chapter 1 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

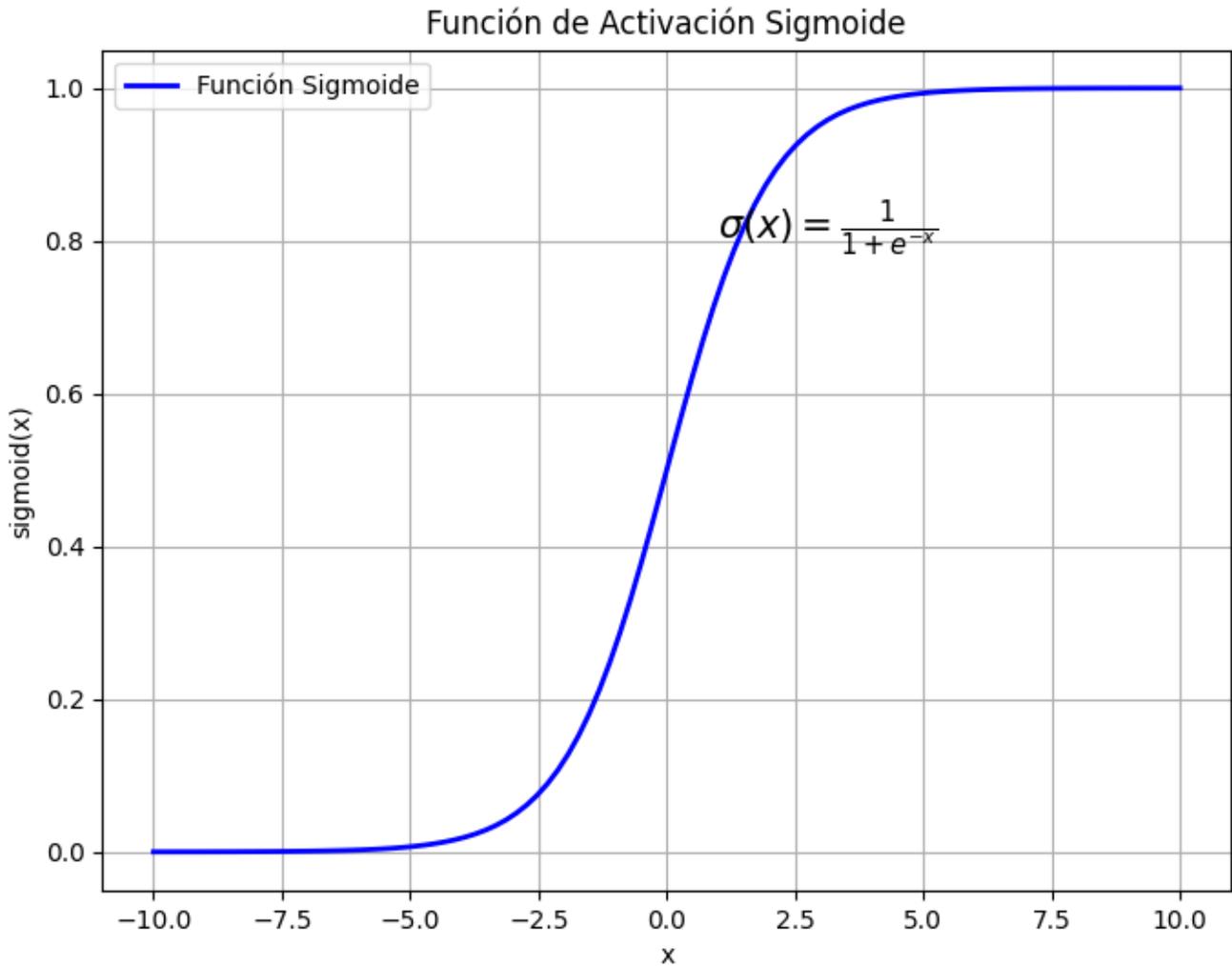


Figura 6. Función sigmoide

En donde en general para una neurona la ecuación correspondiente a la activación se muestra en la ecuación 3:

$$a_0^1 = \sigma(w_{0,0}a_0^0 + w_{0,1}a_1^0 + w_{0,2}a_2^0 + \dots + w_{0,n}a_n^0 + b_0) \quad EC.(3)$$

a_0^1 , 1 no es un exponente es un superíndice que nos indica el número de la capa en la que nos encontramos y 0 es un subíndice que nos indica la neurona en la que estamos.

$w_{j,k}$, puesto que w es un peso, es decir una conexión, j representa la conexión en la neurona a la que estamos situados y k es la conexión que cierra dicha conexión con otra neurona. En la

siguiente matriz podemos observar las operaciones matemáticas que serán la entrada a nuestra función de activación como se muestra en la ecuación 4.

$$\left(\begin{bmatrix} W_{0,0} & W_{0,1} & \dots & W_{0,n} \\ W_{1,0} & W_{1,1} & & W_{1,n} \\ \vdots & & \ddots & \vdots \\ W_{k,0} & W_{0,k} & \dots & W_{k,n} \end{bmatrix} \begin{bmatrix} a_0^0 \\ a_1^0 \\ \vdots \\ a_n^0 \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ n \end{bmatrix} \right) \text{ EC. (4)}$$

Imaginemos que poseemos una imagen para realizar un reconocimiento inteligente mediante nuestra red neuronal. La dimensión de la imagen es de 10,000 píxeles, y contamos con dos subcapas compuestas por 20 neuronas cada una, junto con una capa de salida que alberga 5 neuronas. El número total de pesos a ajustar en nuestra red es de $10,000 \times 20 + 20 \times 20 + 20 \times 5$, resultando en 200,500 pesos. Adicionalmente, se suman $20 + 20 + 5$ sesgos, lo que da un total de 45 parámetros. En total, contamos con 200,545 parámetros que necesitan ser ajustados.

Esta magnitud representa un doble desafío. En primer lugar, la inmensa complejidad para definir cada uno de estos parámetros genera la interrogante de por dónde empezar. En segundo lugar, la red exhibe una sensibilidad notable, ya que cada parámetro definido ejercerá un efecto cascada en las capas subsiguientes. Este panorama resalta la complejidad intrínseca de ajustar y optimizar un gran número de parámetros, planteando interrogantes fundamentales sobre cómo abordar eficazmente esta tarea y cómo garantizar que la red sea robusta ante variaciones en sus parámetros.

Durante el proceso de entrenamiento de una red neuronal, se le suministran datos de entrenamiento previamente etiquetados, lo que significa que la solución ya es conocida de antemano (aprendizaje supervisado). En este escenario, nuestra red aprende a ajustar sus pesos y sesgos para reflejar con precisión los patrones presentes en los datos. Este ajuste se realiza mediante la creación de una función de costos, que compara la salida predicha por la red con los datos etiquetados. En esencia, se suman los cuadrados de las diferencias entre la

salida de la capa final y el valor de entrenamiento esperado. Este proceso es iterativo se realiza individualmente con cada dato de entrenamiento, y su objetivo es reducir gradualmente el costo, buscando alcanzar el mínimo de dicha función de costos¹³.

Una red neuronal se define por un número determinado de entradas y una capa de salida, y se caracteriza por la parametrización de sus miles de pesos y sesgos. La función de costos, que toma estos parámetros como entrada, genera un solo número que indica la efectividad de la red en su capacidad para modelar los datos de entrenamiento. Esta función de costos se parametriza a través de numerosos ejemplos de entrenamiento, refinando así el desempeño de la red neuronal a medida que progresa en su proceso de aprendizaje¹⁴.

Indudablemente, la función de costos no solo nos proporciona información sobre el desempeño de la red, sino que también sirve como el mecanismo mediante el cual indicamos a la red neuronal cómo ajustar los miles de pesos y sesgos que la conforman. Este ajuste se logra mediante la aplicación del descenso del gradiente¹⁵, una técnica que busca encontrar un mínimo local de la función de costos, dado que encontrar el mínimo global es una tarea ardua y desafiante puesto que la función de costos es una función con miles de dimensiones. El descenso del gradiente se expresa con un signo negativo con sus respectivas derivadas parciales como se muestra a continuación en la ecuación.

¹³ 3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

¹⁴ 3Blue1Brown." Neural Networks, Chapter 2 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

¹⁵ 3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

$$-\nabla C(\vec{W}) = \begin{bmatrix} \frac{\partial C}{\partial w^{(1)}} \\ \frac{\partial C}{\partial b^{(1)}} \\ \vdots \\ \frac{\partial C}{\partial w^{(L)}} \\ \frac{\partial C}{\partial b^{(L)}} \end{bmatrix} \quad EC. (5)$$

El vector (\vec{W}) representa una colección de pesos y sesgos de nuestro modelo, mientras que el gradiente negativo nos proporciona la dirección, expresada como un vector, en la cual la función de coste disminuirá. En este contexto, se hace necesario actualizar este vector y repetir el proceso iterativo hasta alcanzar el óptimo deseado. Las derivadas parciales, ecuaciones que abordaré más adelante, desempeñan un papel crucial en este proceso de optimización.

Una vez que nuestra red esté completamente entrenada y su eficacia alcance niveles significativamente altos, alcanzando un 94% o incluso un 97% de precisión, es probable que se encuentre en un mínimo local en el que identifique patrones¹⁶. Sin embargo, es importante señalar que estos patrones identificados pueden no ser necesariamente los esperados. En otras palabras, nuestra red puede reconocer ciertos atributos sin comprender plenamente qué atributos son exactamente. Este fenómeno resalta la complejidad inherente del proceso de aprendizaje automático y la necesidad de una evaluación cuidadosa y continua para comprender y corregir posibles desviaciones en el rendimiento del modelo.

¿Cuáles pesos y sesgos minimizan la función de costos? Cada componente del vector de gradiente aporta información crucial, revelándonos qué cambios son más sensibles en términos de magnitud. En este contexto, no es igual generar modificaciones en una componente que en otra. El descenso de gradiente se basa en cada dato de entrenamiento, donde todos los

¹⁶ 3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

gradientes negativos de cada entrenamiento se suman y promedian al final. Estos ajustes dependen de la distancia que separa los valores actuales de los objetivos.

Ahora bien, consideremos una subcapa donde deseamos aumentar la activación de un perceptrón. Recordemos que esta activación está influenciada por las activaciones de la capa anterior, sumadas al sesgo, y todo esto encapsulado en una función de activación no lineal. Para modificar el nivel de activación, tenemos tres enfoques: aumentar el sesgo, aumentar el peso o cambiar las activaciones de la capa anterior indirectamente. En cuanto a los pesos, estos pueden estar conectados a activaciones tanto altas como bajas. Por lo tanto, ajustar el valor de un peso asociado a una activación alta tendrá un impacto significativo en la función de costos¹⁷. Es fundamental comprender que el vector de gradiente no solo nos indica los cambios necesarios en los pesos y sesgos, sino que también señala cuáles son los más relevantes.

Otra estrategia es modificar las activaciones de la capa anterior en función de la intensidad del peso conectado. Si el peso es bajo, será óptimo reducir el incremento de la activación precedente. Sin embargo, no podemos alterar directamente esta activación; solo podemos hacerlo a través de los pesos y sesgos correspondientes¹⁸. Todos estos efectos se acumulan y se promedian para la penúltima capa, repitiendo este proceso recursivamente mientras retrocedemos en la red. Este meticuloso procedimiento asegura que la red se ajuste de manera adecuada, capturando de forma precisa los patrones en los datos de entrenamiento.

Comencemos a comprender los cálculos. Para este ejemplo, necesitaremos una red simple.

¹⁷ 3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

¹⁸ 3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

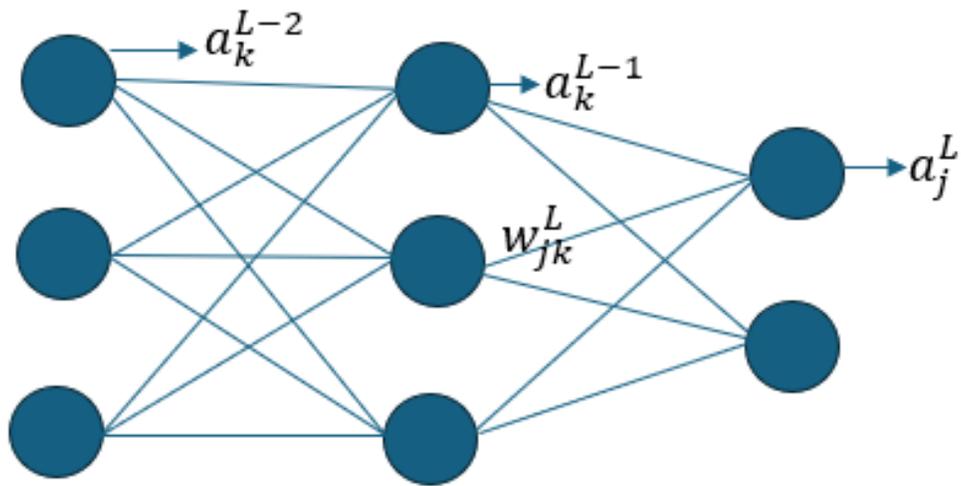


Figura 7. Red neuronal simple

Imagen inspirada de los artículos de 3blue1brown

Referencia: (3blue1brown, 2017)

<https://www.3blue1brown.com/topics/neural-networks>

Suponiendo que nuestro dato de entrenamiento sea y_j , por lo tanto, el costo de esta red para un ejemplo de entrenamiento dado es:

$$C(a_j^L) = \frac{1}{2} \sum_{j=0}^{nL-1} (a_j^L - y_j)^2 \quad \text{EC. (6)}$$

- y_j Es la predicción de la red para el ejemplo j .
- a_j^L es el valor real para el ejemplo j , L no es un exponente es un superíndice que indica que está en la última capa para las neuronas que no estén en la última capa su nomenclatura es a_k^L

Necesitamos saber cómo varía nuestra función de costo para cada parámetro de la red, para eso necesitamos derivar parcialmente obteniendo las siguientes ecuaciones:

Nos ubicamos en la última capa para deducir las siguientes ecuaciones, tomaremos una única conexión entre una neurona de la última capa con otra neurona de la capa anterior, obteniendo:

$$a_k^{L-1} w_{jk}^L + b_j^L = z_j^L \text{ EC. (7)}$$

Aquí podemos observar que simplificamos nuestra suma ponderada con una la nueva variable z_j^L , por lo tanto, nuestra actuación en la última capa está dada por la siguiente ecuación:

$$a_j^L = \sigma(z_j^L) \text{ EC. (8)}$$

Para saber cómo varía nuestra función de costos respecto a los pesos utilizamos la regla de la cadena puesto que nuestra función de costos no está en términos de w_{jk}^L

$$\frac{\partial C_0}{\partial w_{jk}^L} = \frac{\partial z_j^L}{\partial w_{jk}^L} \frac{\partial a_j^L}{\partial z_j^L} \frac{\partial C_0}{\partial a_j^L} \text{ EC. (9)}$$

Calculando las derivadas parciales:

$$\frac{\partial C_0}{\partial a_j^L} = (a_j^L - y_j) \text{ EC. (10)}$$

$$\text{Si } a_j^L(z_j^L) = \frac{1}{1 + e^{-z_j^L}} \text{ EC. (11)}$$

$$\frac{\partial a_j^L}{\partial z_j^L} = a_j^L(z_j^L)(1 - a_j^L(z_j^L)) \text{ EC. (12)}$$

$$\text{Si } z_j^L = \sum_j a_k^{L-1} w_j^L + b_j^L \text{ EC. (13)}$$

$$\frac{\partial z_j^L}{\partial w_j^L} = a_k^{L-1} \text{ EC. (14)}$$

En esta derivada podemos concluir que la sensibilidad en nuestra suma ponderada cuando hay cambios en el peso dependen de la intensidad de la activación anterior.

Ahora calculando la derivada de la función de costo respecto al sesgo:

$$\frac{\delta C_0}{\partial b_j^L} = \frac{\partial z_j^L}{\partial b_j^L} \frac{a_j^L}{\partial z_j^L} \frac{\partial C_0}{\partial a_j^L} EC. (15)$$

$$\frac{\partial z_j^L}{\partial b_j^L} = 1 EC. (16)$$

Puesto que estamos situado en la última capa, ya calculamos nuestras derivadas respecto a nuestros parámetros $w_j^L + b_j^L$, ahora calculando la derivada respecto la activación anterior:

$$\frac{\partial C_0}{\partial a_k^{L-1}} = \frac{\partial z_j^L}{\partial a_k^{L-1}} \frac{a_j^L}{\partial z_j^L} \frac{\partial C_0}{\partial a_j^L} EC. (17)$$

$$\frac{\partial z_j^L}{\partial a_k^{L-1}} = w_{jk}^L EC. (18)$$

Puesto que nuestra activación anterior puede influir en distintas neuronas de nuestra última capa se obtiene:

$$\frac{\partial C_0}{\partial a^{L-1}} = \sum_{j=0}^{nL-1} \frac{\partial z_j^L}{\partial a^{L-1}} \frac{\partial a_j^L}{\partial z_j^L} \frac{\partial C_0}{\partial a_j^L} = w_{jk}^L EC. (19)$$

El resultado de esta derivada representa la matriz de pesos en las que están conectadas dos capas.

Todos estos cálculos fueron para la última capa ahora retrocedemos en la red obteniendo las siguientes ecuaciones:

$$\frac{\partial C_0}{\partial w_{jk}^{L-1}} = \frac{\partial z_j^{L-1}}{\partial w_{jk}^{L-1}} \frac{\partial a_j^{L-1}}{\partial z_j^{L-1}} \frac{\partial z_j^L}{\partial a_j^{L-1}} \frac{\partial a_j^L}{\partial z_j^L} \frac{\partial C_0}{\partial a_j^L} EC. (20)$$

$$\frac{\partial C_0}{\partial b_j^{L-1}} = \frac{\partial z_j^{L-1}}{\partial b_j^{L-1}} \frac{\partial a_j^{L-1}}{\partial z_j^{L-1}} \frac{\partial z^L}{\partial a_j^{L-1}} \frac{\partial a_j^L}{\partial z^L} \frac{\partial C_0}{\partial a_j^L} \text{ EC. (21)}$$

Analizando esta ecuación podemos ver que $\frac{\partial a_j^L}{\partial z^L} \frac{\partial C_0}{\partial a_j^L}$ y $\frac{\partial a_j^L}{\partial z^L} \frac{\partial C_0}{\partial a_j^L}$ son ecuaciones que ya calculamos en nuestra capa anterior, lo que simplifica los cálculos, $\frac{\partial a_j^{L-1}}{\partial z_j^{L-1}}$ y $\frac{\partial a_j^{L-1}}{\partial z_j^{L-1}}$ son las derivadas respecto a la función de activación, por último $\frac{\partial z_j^{L-1}}{\partial w_{jk}^{L-1}}$ y $\frac{\partial z_j^{L-1}}{\partial b_j^{L-1}}$ son derivadas respecto a los pesos y sesgos que se calculan de la misma manera que en la capa anterior, por último el cálculo de $\frac{\partial z^L}{\partial a_j^{L-1}}$ y $\frac{\partial z^L}{\partial a_j^{L-1}}$ nos permitirá poder conocer nuestra matriz de pesos que conectan ambas capas. Es importante mencionar que todas las ecuaciones mencionadas en el escrito fueron obtenidas del libro "Neural Networks and Deep Learning: Introduction to the Core Principles" de Michael Nielsen. Este libro es una excelente fuente de información para comprender los fundamentos de las redes neuronales y el aprendizaje profundo.

Este algoritmo es conocido como el algoritmo de retropropagación, en el cual calculamos las derivadas de nuestra función de costo con respecto a cada parámetro de la red¹⁹. Esto nos permite optimizar nuestra función y comprender el error de la red en cada subcapa y parámetro. No importa cuántas neuronas o subcapas tenga la red, estas ecuaciones nos permitirán conocer las variaciones de cada parámetro para modificarlos y repetir este proceso hasta minimizar la función de costos. Como resultado final, obtendremos una red entrenada que se comportará de manera inteligente, logrando reconocer patrones de manera efectiva.

3. ANÁLISIS DEL PROBLEMA Y SOLUCIÓN IMPLEMENTADA

GEICO, una destacada empresa de seguros se encuentra ante el desafío de modernizar y agilizar sus procesos de reclamaciones en diversas ramas de negocio, que incluyen seguros de automóviles, motos, viviendas, salud y empresas. En su búsqueda de eficiencia y mejora en la

¹⁹ 3Blue1Brown." Neural Networks, Chapter 4 - 3Blue1Brown, <https://www.3blue1brown.com/lessons/neural-networks>.

atención al cliente, GEICO ha proporcionado una variedad de documentos relacionados con reclamaciones, como recibos, tickets, facturas digitales, facturas físicas, PDF de informes de siniestros, informes de laboratorio médico, clínico, rehabilitación etc. Cada tipo de negocio conlleva su propio conjunto de documentos, que deben ser procesados de manera eficaz y precisa para gestionar las reclamaciones de manera oportuna. La diversidad de formatos y tipos de documentos presenta un desafío que requiere la implementación de soluciones de automatización inteligente para acelerar la gestión de reclamaciones y mejorar la experiencia del cliente.

Para compartir de manera segura la información necesaria con Shift Technology para el proyecto de Automation Claim, GEICO ha empleado prácticas de seguridad informática rigurosas. Primero, se ha utilizado una conexión cifrada mediante protocolos seguros, como HTTPS²⁰ o SFTP²¹, para garantizar la confidencialidad de los datos durante su transferencia. Además, GEICO ha implementado medidas de autenticación fuerte, como la autenticación de dos factores, para verificar la identidad de los usuarios que acceden a la información.

La información se ha compartido en formatos de archivo seguros. Se han aplicado controles de acceso estrictos para limitar el acceso solo a personal autorizado de Shift Technology y GEICO. La comunicación entre ambas partes se ha realizado a través de canales de comunicación seguros, y se han establecido acuerdos formales de confidencialidad y seguridad de datos para garantizar el cumplimiento de normativas y la protección de la información.

²⁰ Mozilla. "Hypertext Transfer Protocol Secure (HTTPS)." Mozilla Developer Network. Recuperado de: <https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview>.

²¹ VanDyke Software. "What is SFTP (SSH File Transfer Protocol)?" VanDyke Software. Recuperado de: <https://www.vandyke.com/products/securefx/sftp.html>.

3.1 Análisis del problema

El dataset de estudio lo generé utilizando técnicas de web scraping²², utilizando bibliotecas populares como BeautifulSoup²³, Selenium²⁴, PIL²⁵, Tesseract²⁶ y Requests²⁷ en python, en las cuales el equipo de IT(Infrastructure Team) creó un servidor con protocolos de seguridad adecuados para acceder a la información de diversas fuentes en línea de manera automatizada y controlada. Este proceso me permitió obtener un conjunto de datos que consta de 13 mil documentos, que incluyen una variedad de tipos de contenido, como texto, imágenes, PDF, y datos tabulares. La diversidad de formatos y tipos de documentos en nuestro dataset representa un gran desafío puesto que se tiene que procesar dicha información en información cifrable para poder utilizar técnicas posteriores en su correspondiente clasificación. En la figura 8 se muestra el proceso de generación del dataset.

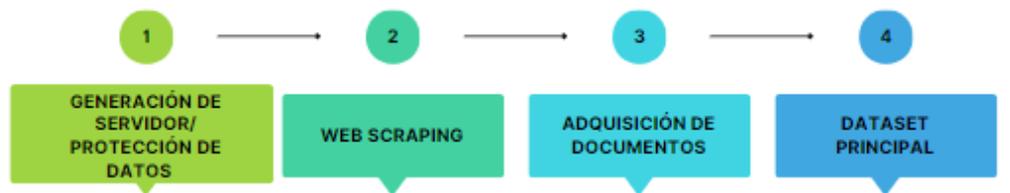


Figura 8. Metodología de generación de data set

3.2 Solución implementada

La generación de nuestro dataset nos brinda un punto de partida esencial para llevar a cabo un análisis detallado de los datos recopilados. Con estos 13 mil documentos a nuestra disposición, podemos comenzar a explorar su contenido y estructura, lo que nos permitirá entender mejor la naturaleza de la información que contienen y cómo se presenta. Después de la generación del

²² Subirats, L., & Calvo, M. (2018). "Web Scraping." O'Reilly Media. <https://www.oreilly.com/library/view/web-scraping-with/9781491985564/>.

²³ Richardson, L. (2019). "Beautiful Soup Documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.

²⁴ SeleniumHQ. "Selenium Documentation, <https://www.selenium.dev/documentation/en/>.

²⁵ PythonWare. (2011). "Python Imaging Library Handbook.", <https://pillow.readthedocs.io/en/stable/handbook/index.html>.

²⁶ Tesseract OCR. "Tesseract Documentation.", <https://github.com/tesseract-ocr/tesseract/wiki>.

²⁷ The Python Requests Team. "Requests Documentation.", <https://docs.python-requests.org/en/latest/>.

dataset, se empleó la técnica de OCR (Optical Character Recognition)²⁸ para llevar a cabo el reconocimiento de las palabras clave contenidas en los documentos. Esta fase de procesamiento permitió convertir las imágenes y texto escaneado en formato digital legible por la máquina, lo que facilitó su posterior análisis y extracción de información relevante, obteniendo la siguiente información.

Realicé una gráfica de frecuencia para poder analizar el dataset a simple vista obteniendo los siguientes resultados que se muestran en la figura 9.

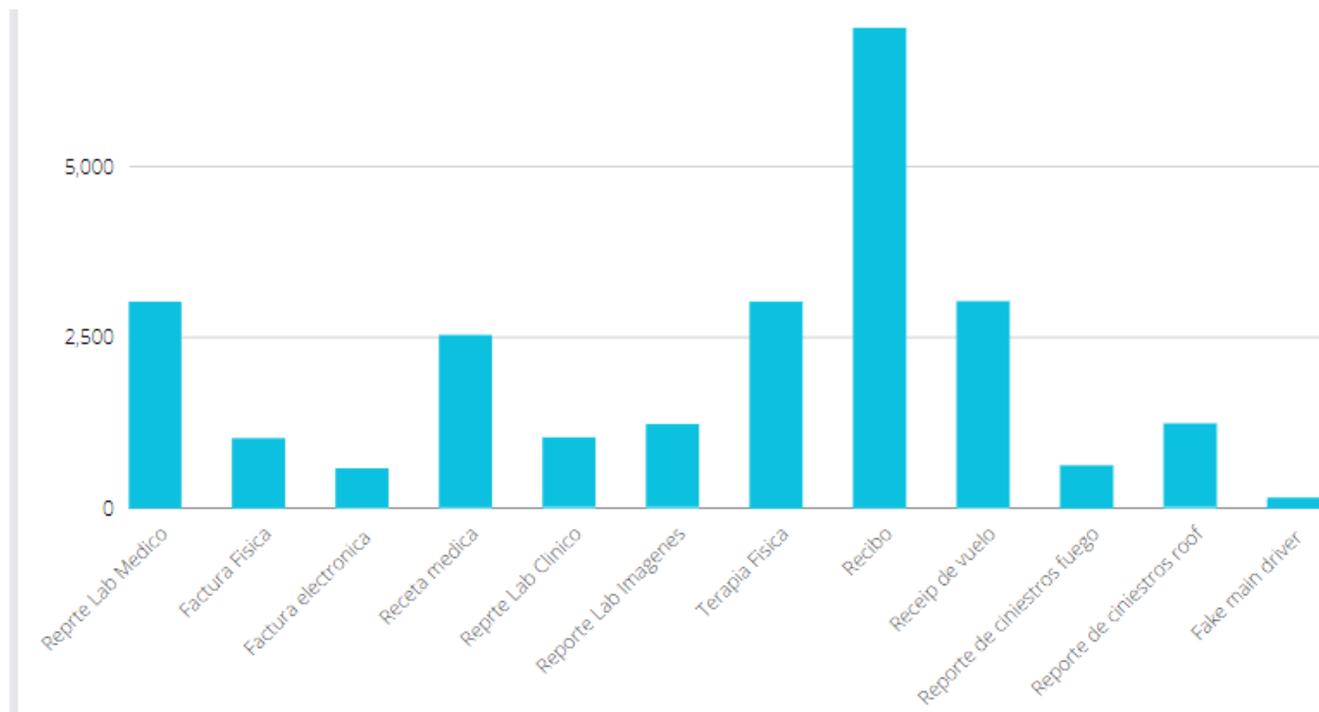


Figura 9. Tabla de frecuencia

Tras aplicar la técnica de OCR, llevé a cabo una inspección visual detallada de los resultados obtenidos. Durante este proceso, se identificaron y descartaron las palabras mal detectadas por el modelo OCR, asegurando así la precisión y calidad de los datos procesados. Se implementó

²⁸ Smith, R. (2007). "An Overview of the Tesseract OCR Engine." Proceedings of the Ninth International Conference on Document Analysis and Recognition. Recuperado de: <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/33418.pdf>.

un enfoque meticuloso para garantizar la calidad y la idoneidad de los datos para el entrenamiento del modelo. Esta revisión manual fue crucial para garantizar la integridad y confiabilidad del dataset, preparándolo para análisis y aplicaciones posteriores. Este proceso involucró la revisión cuidadosa de cada documento para identificar y eliminar aquellos con calidad deficiente, contenido irrelevante o información duplicada. Además, se aplicaron técnicas de Análisis Exploratorio de Datos (EDA)²⁹ para limpiar y preprocesar los datos de manera efectiva. Esto incluyó la detección y eliminación de errores de formato, la normalización del texto y la corrección de cualquier inconsistencia o ambigüedad en los datos.

3.3 Metodología

Solicité al equipo de infraestructura (IT) el desarrollo de una interfaz gráfica denominada "Stormtrooper". En esta interfaz, especifiqué atributos particulares que permitirían la categorización efectiva de los documentos utilizando etiquetas digitales. Estos atributos fueron cuidadosamente diseñados para capturar información relevante, facilitando así la organización y búsqueda eficiente en el dataset. La interfaz "Stormtrooper" proporcionó una plataforma intuitiva para asignar categorías a los documentos, optimizando así el proceso de categorización y clasificación.

En el proceso de entender y categorizar los datos de entrenamiento se proporcionaron categorizaciones específicas que resultaron esenciales para subcategorizar los documentos de manera precisa. Por ejemplo, para los tickets, se detallaron las etiquetas necesarias para desglosar los precios, mientras que para los recibos de vuelos se incluyeron etiquetas como el número de pasajeros, así como las horas de salida y llegada del vuelo. Este enfoque detallado en la categorización de datos permitió un entrenamiento más profundo y preciso de nuestro modelo de machine learning, asegurando que el sistema pudiera entender y procesar de manera efectiva la información específica contenida en diversos tipos de documentos. Estas etiquetas proporcionaron una guía valiosa para el algoritmo, permitiéndonos desarrollar soluciones de procesamiento documental que se adaptaran y respondieron de manera inteligente a las necesidades variadas de los documentos en nuestro conjunto de datos. Normalmente, este tipo de subcategorías no es muy común, pero recordemos que este proyecto

²⁹ Tufte, E. R. (2001). "The Visual Display of Quantitative Information." Graphics Press.

no solo es para la clasificación de documentos, sino también para comprender esos documentos y permitir que nuestros algoritmos aprendan a identificar fraudes. Por eso, es importante que nuestra red artificial pueda identificar todos los atributos del documento correspondiente. A continuación, en las figuras 10-13 se muestran las siguientes categorías de nuestros documentos.



Figura 10. Categorización de Reporte Laboratorio Procesos

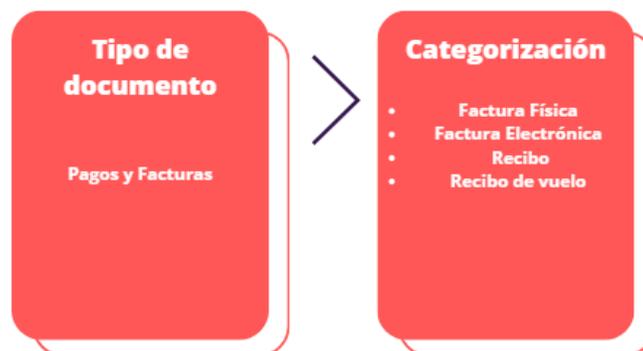


Figura 11. Categorización de Pagos y Facturas

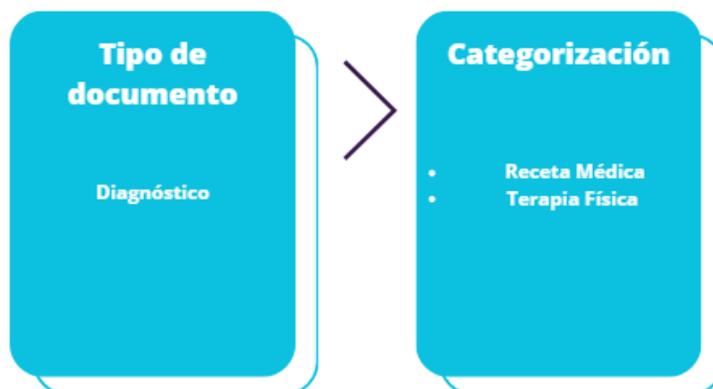


Figura 12. Categorización Diagnóstico



Figura 13. Categorización Siniestros

Para esta tarea de clasificación de imágenes, utilicé una red neuronal convolucional, que es especialmente efectiva para procesar datos de imágenes. La arquitectura de la red consistió en varias capas convolucionales seguidas de capas de pooling para extraer características de las imágenes. Luego, agregué capas totalmente conectadas para realizar la clasificación final.

En términos de subcapas, utilicé una combinación de capas convolucionales 2D y capas de pooling, seguidas de capas totalmente conectadas. La cantidad de subcapas depende del tamaño, utilicé alrededor de 3-4 capas convolucionales seguidas de capas de pooling³⁰ y 1-2 capas totalmente conectadas al final.

Para evitar el sobreajuste, implementé la técnica de Dropout en algunas de las capas totalmente conectadas. El Dropout ayuda a regularizar la red neuronal al apagar aleatoriamente un cierto

³⁰ Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning." MIT Press. Chapter 9: Convolutional Networks, pp. 315-359.

porcentaje de neuronas durante el entrenamiento, lo que evita la dependencia excesiva de ciertas características y mejora la generalización del modelo³¹.

Utilicé la función de pérdida de entropía cruzada categórica, ya que estamos realizando una tarea de clasificación multiclase. Esta función de pérdida es adecuada para problemas de clasificación y mide la discrepancia entre las distribuciones de probabilidad reales y predichas. Dividí el conjunto de datos en lotes de tamaño apropiado para el entrenamiento de la red. El tamaño del lote dependió del tamaño del conjunto de datos, pero en general, utilicé lotes de alrededor de 130-150 imágenes. Realicé múltiples épocas de entrenamiento para permitir que la red neuronal aprendiera de manera efectiva las características de las imágenes. El número específico de épocas de entrenamiento varió según la complejidad del problema y la convergencia del modelo, pero en general, utilicé alrededor de 100 a 150 épocas, posteriormente implementé la técnica de Early Stopping³² para detener el entrenamiento si el rendimiento del modelo en el conjunto de validación dejaba de mejorar. Esta técnica ayuda a evitar el sobreajuste al detener el entrenamiento una vez que el rendimiento del modelo comienza a empeorar en el conjunto de validación.

4. RESULTADOS

Después de analizar los resultados del entrenamiento de la red neuronal, se observa que en la época 93 nuestra red neuronal alcanza una eficiencia del 95%, lo que indica un rendimiento bastante sólido en la clasificación de las imágenes como se muestra en la figura 14. Sin embargo, es importante tener en cuenta que el comportamiento de la red posterior a esta época podría variar.

Es posible que la eficiencia continúe aumentando gradualmente en las épocas posteriores, aunque a un ritmo más lento debido a la convergencia del modelo y la disminución de la tasa de mejora. También existe la posibilidad de que la eficiencia se estabilice o incluso disminuya

³¹ Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning." MIT Press. Capítulo 7.12: Dropout, pp. 248-251.

³² Bishop, C. M. (2006). "Pattern Recognition and Machine Learning." Springer. Capítulo 7.10: Early Stopping, pp. 222-224.

ligeramente debido al sobreajuste del modelo a los datos de entrenamiento, especialmente si no se aplican técnicas de regularización adecuadas.

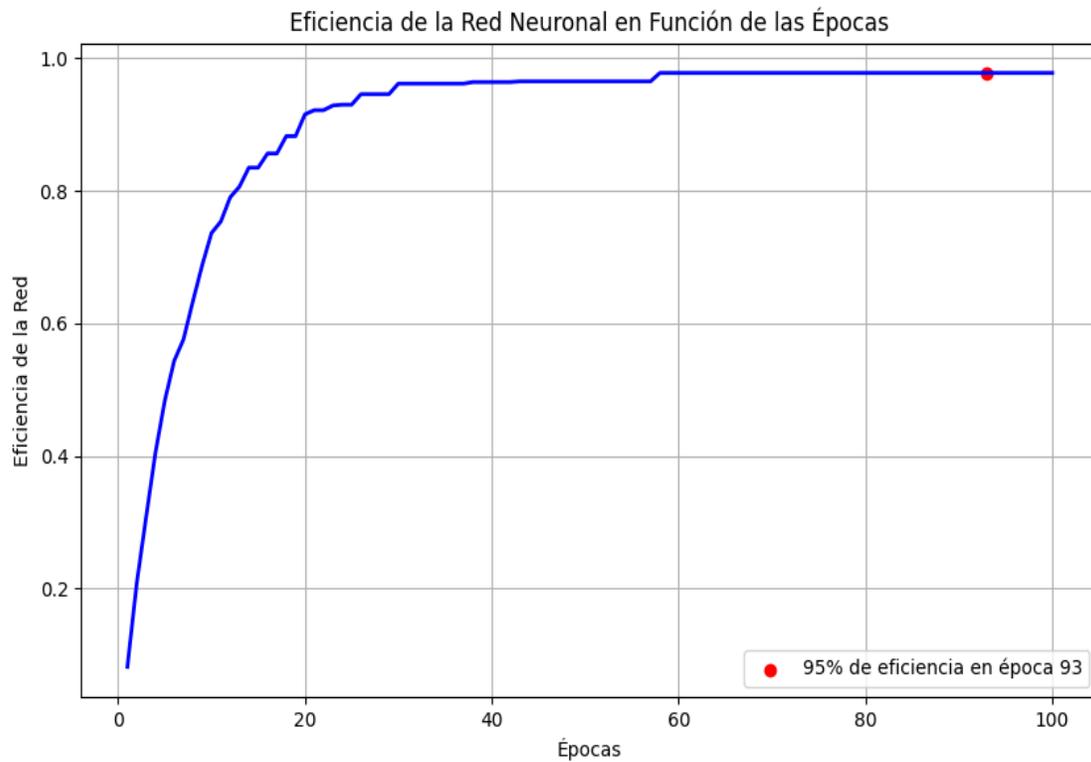


Figura 14. Eficiencia de la red

Además, otros resultados que podríamos mencionar incluyen la tendencia general de mejora de la eficiencia a lo largo de las épocas, la variabilidad en los valores de eficiencia debido a las fluctuaciones aleatorias y la importancia de monitorear el rendimiento del modelo en conjuntos de datos de validación para evitar el sobreajuste.

5. CONCLUSIONES

El proceso de entrenamiento se ejecutó de manera exitosa, logrando etiquetar los 13 mil documentos con sus subetiquetas correspondientes en un tiempo récord de 3 semanas. Este proceso de etiquetado fue llevado a cabo por el equipo de Q&A Data Scientists, del cual formé parte activa. Durante estas intensas tres semanas, trabajamos incansablemente para asegurar que cada documento fuera etiquetado con precisión y coherencia, permitiendo así que nuestra red neuronal pudiera aprender patrones y entender la complejidad de los datos de manera efectiva.

En conclusión, este proyecto me ha servido como una valiosa manera de explicar un breve introducción al funcionamiento del machine learning, centrándose en el uso de redes neuronales para la clasificación de imágenes. A través de este proceso, he podido comprender mejor mi rol y mi experiencia como Data Scientist, adquiriendo conocimientos prácticos sobre cómo diseñar, entrenar y evaluar modelos de aprendizaje automático para resolver problemas específicos. Si bien mi intención no es adentrarme en detalles altamente técnicos ni en ecuaciones más complejas, sino proporcionar los elementos necesarios para comprender los fundamentos de la inteligencia artificial y narrar mi experiencia como Q&A Data Scientist, es importante señalar que se han omitido numerosos aspectos. Esto se ha hecho con el propósito de no abrumar al lector con temas que están estrechamente relacionados y que implican un nivel de dificultad más avanzado, los cuales quedan fuera del alcance de este informe.

Agradezco sinceramente a la Facultad de Ingeniería por proporcionarme una sólida formación académica y las bases suficientes para entender y aplicar tecnologías emergentes como el machine learning y la Inteligencia Artificial. Los conocimientos y habilidades adquiridos durante mi formación académica me han permitido abordar este proyecto con confianza y capacidad para enfrentar los desafíos que surgen en el campo del Data Science.

Este proyecto también ha reforzado mi comprensión sobre la importancia del aprendizaje continuo y la adaptabilidad en un campo tan dinámico como el Data Science. A medida que avanzo en mi carrera profesional, estoy comprometido a seguir explorando nuevas tecnologías,

mejorando mis habilidades y contribuyendo de manera significativa al desarrollo y la innovación en el campo del análisis de datos y la Inteligencia Artificial.

Este proyecto ha sido una experiencia enriquecedora que ha ampliado mi comprensión sobre el machine learning y ha reforzado mi pasión por la ciencia de datos. Estoy emocionado por las oportunidades futuras que me esperan y estoy agradecido por todo el apoyo y los recursos que he recibido durante mi formación académica en la Facultad de Ingeniería.

BIBLIOGRAFÍA Y REFERENCIAS

1. Shift Technology.(2021). Series D Investment Round Led by Advent International. Obtenido de <https://www.shift-technology.com/resources/news/shift-technology-secures-series-d#:~:text=PARIS%20and%20BOSTON%20%E2%80%93%20May%2006,investment%20round%20of%20%24220%20million.>
2. SVM Tutorial. (2014) SVM - Understanding the math, Obtenido de <https://www.svm-tutorial.com/>
3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." Nature, 521(7553), 436-444.
4. Monkeylearn. (2028) Naive Bayes Classifier Explained, Obtenido de <https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/>
5. Towardsdatascience (2019), Logistic Regression Explained, Obtenido de <https://towardsdatascience.com/understanding-logistic-regression-9b02c2aec102>
6. Towardsdatascience (2019) ,Understanding Random Forests, Obtenido de <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
7. Divason, J., Martinez De Pison, F.J., Romero, A., & Saenz-De-Cabezón, E. (2023). "Artificial Intelligence Models for Assessing the Evaluation Process of Complex Student Projects." IEEE Transactions on Learning Technologies.
8. Mitchell, T. M. (1997). "Machine Learning." McGraw Hill, New York.
9. Bishop, C. M. (2006). "Pattern Recognition and Machine Learning." Springer.
10. Sutton, R. S., & Barto, A. G. (2018). "Reinforcement Learning: An Introduction." The MIT Press.

11. Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). "Semi-Supervised Learning." The MIT Press.
12. "3Blue1Brown." Neural Networks, Chapter 1 - 3Blue1Brown, Obtenido de <https://www.3blue1brown.com/lessons/neural-networks>.
13. "3Blue1Brown." Neural Networks, Chapter 2 - 3Blue1Brown, Obtenido de <https://www.3blue1brown.com/lessons/neural-networks>.
14. "3Blue1Brown." Neural Networks, Chapter 3 - 3Blue1Brown, Obtenido de <https://www.3blue1brown.com/lessons/neural-networks>.
15. "3Blue1Brown." Neural Networks, Chapter 4 - 3Blue1Brown, Obtenido de <https://www.3blue1brown.com/lessons/neural-networks>.
16. Mozilla. "Hypertext Transfer Protocol Secure (HTTPS)." Mozilla Developer Network. Obtenido <https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview>.
17. VanDyke Software. "What is SFTP (SSH File Transfer Protocol)?" VanDyke Software. Obtenido de: <https://www.vandyke.com/products/securefx/sftp.html>.
18. Subirats, L., & Calvo, M. (2018). "Web Scraping." O'Reilly Media. <https://www.oreilly.com/library/view/web-scraping-with/9781491985564/>.
19. Richardson, L. (2019). "Beautiful Soup Documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
20. SeleniumHQ.(2023). "Selenium Documentation, Obtenido de <https://www.selenium.dev/documentation/en/>.
21. PythonWare. (2011). "Python Imaging Library Handbook.", Obtenido de <https://pillow.readthedocs.io/en/stable/handbook/index.html>.
22. Tesseract OCR. "Tesseract Documentation.", Obtenido de <https://github.com/tesseract-ocr/tesseract/wiki>.
23. The Python Requests Team. "Requests Documentation.", <https://docs.python-requests.org/en/latest/>.
24. Smith, R. (2007). "An Overview of the Tesseract OCR Engine." Proceedings of the Ninth International Conference on Document Analysis and Recognition. Obtenido de <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/33418.pdf>.

25. Tufte, E. R. (2001). "The Visual Display of Quantitative Information." Graphics Press.
26. Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning." MIT Press. Chapter 9: Convolutional Networks, pp. 315-359.
27. Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning." MIT Press. Capítulo 7.12: Dropout, pp. 248-251.
28. Bishop, C. M. (2006). "Pattern Recognition and Machine Learning." Springer. Capítulo 7.10: Early Stopping, pp. 222-224.