



TRABAJO DE GRADO
Opción Seminario-Diplomado.

Proyecto Machine Learning de predicción de pedidos para bodega de almacenamiento
Seminario Machine Learning e Inteligencia Artificial

Corporación Universitaria Remington.
Ingeniería de Sistemas

Víctor Manuel Hernández Vargas

Docente Juan Pablo Vélez Uribe
Opción de Trabajo de grado Seminario.
2023

Agradecimientos

Feliz de completar el proceso de formación, quiero elaborar este trabajo con dedicatoria a la institución Uniremington por el apoyo brindado durante estos años académicos, a todo el equipo de docentes, que, aunque a distancia, siempre estuvieron disponibles para apoyar el proyecto formativo, instruyendo, evaluando, apoyando y sobre todo acompañando.

Agradecimiento especial a mi familia por el apoyo brindado durante estos años de formación académica.

Tabla de Contenidos

Resumen.....	4
Marco conceptual y contextual	5
1. Introducción al Machine Learning.....	5
2. Fundamentos aplicados de Machine Learning.....	8
3. Introducción a la ética en la inteligencia artificial	10
4. Evaluación de modelos de Machine Learning	11
5. Introducción a la Inteligencia Artificial	16
6. Machine Learning: Análisis contrafactual	18
7. Machine Learning: Aprendizaje supervisado	20
8. Innovación tecnológica con inteligencia artificial	21
Desarrollo e implementación del aprendizaje.....	22
Figuras y tablas	28
Conclusiones.....	29
Bibliografía	30

Resumen

Este proyecto se basa en el seminario de Machine Learning e Inteligencia Artificial ofrecido por la plataforma Crehana compuesto por 8 módulos

- Introducción al Machine Learning
- Fundamentos aplicados de Machine Learning
- Introducción a la ética en la inteligencia artificial
- Evaluación de modelos de Machine Learning
- Introducción a la Inteligencia Artificial
- Machine Learning: Análisis contrafactual
- Machine Learning: Aprendizaje supervisado
- Innovación tecnológica con inteligencia artificial

En cada uno de ellos se abordan temas de actualidad y allí se adquieren los conocimientos necesarios relacionados al aprendizaje predictivo de las maquinas, así como los diferentes tipos de datos, diferentes tipos de modelos y diferentes tecnologías de desarrollo, fundamentos éticos de la inteligencia artificial y sus diferentes alcances.

En este proyecto se plantea una necesidad a solucionar para la compañía para la cual me encuentro trabajando, la necesidad surge en el abastecimiento de la bodega principal, este abastecimiento actualmente no se realiza de manera óptima, por lo que se sugiere un modelo de clasificación que permita identificar cuando un producto se debe solicitar por compra a un proveedor y por medio de un modelo de regresión, estimar la cantidad a pedir.

Palabras clave

Machine learning, Ejercicios de regresión y Clasificación, Red neuronal, Ética de la inteligencia artificial.

Marco conceptual y contextual

1. Introducción al Machine Learning

Día a día el machine learning nos acompaña en las actividades cotidianas que tenemos al interactuar con las aplicaciones móviles como google, google maps, waze, Facebook, Instagram, spotify, netflix entre otras aplicaciones. [1]

Machine learning está pronosticando, capturando y ejecutando algoritmos que presentan temas de tu interés, películas, series, música, imágenes, publicidad, todo basado en tus comportamientos usuales con las diferentes aplicaciones.

Básicamente el sistema o la maquina “aprende” a mostrarte siempre temas de tu interés con toda la información recolectada, esta información por pequeña o insignificante que parezca está siendo recolectada para el aprendizaje.

Las aplicaciones de machine learning pronostican basándose en datos pasados, siendo la ciencia de datos fundamental en la era actual. Muchas empresas manejan grandes cantidades de datos generados constantemente. Aunque valioso, el uso de datos presenta desafíos, tanto a nivel corporativo como personal. La gestión de información abundante, refleja los retos asociados a su naturaleza. De aquí surge la pregunta clave y es ¿Qué es el machine learning?

Machine learning es un conjunto de herramientas que extraen valor la información almacenada de un sistema, identificando patrones y haciendo predicciones estimando valores, clasificando datos, etc. De esta forma la data se utiliza para convertir datos en inteligencia artificial o básicamente, todos tus datos son usados para el entrenamiento de modelos computacionales, un ejemplo expuesto es el caso de programas de ajedrez donde la maquina aprende de acuerdo a las interacciones del usuario. Por otro lado, el machine learning se entrelaza con el ámbito de Big Data, donde se maneja enormes volúmenes de información, donde la velocidad de procesamiento es crucial y analizar toda la variedad de datos se convierte en una ciencia. Aunque las tecnologías modernas han facilitado su manejo, el machine learning es el nuevo reto a nivel informático para la humanidad, pues se espera que, al entrenar un sistema, este pueda brindar soluciones, cálculos, pronósticos clasificaciones etc. partiendo de los datos históricos previamente almacenados en su data. [2]

En esta exposición se muestra que desafortunadamente no todos los proyectos de machine learning llegan a su éxito o implementación, en muchos casos por que no se tiene un planteamiento de problema claro, o se decide iniciar el proceso invirtiendo en tecnología más que en el análisis de la data existente, se tiene por dato que el 87% de los proyectos de Machine Learning nunca se implementan o salen a su etapa de desarrollo. En conclusión, para toda empresa, o todo proyecto la estrategia del manejo de los datos

es crucial para el éxito en machine learning, Para tener una guía existe una herramienta llamada la pirámide de valor. En ella se definen los puntos clave para poder llevar a cabo la implementación del machine learning y de un proyecto de inteligencia artificial. En esta pirámide la base y quizá lo más importante incluye la tecnología y la gobernanza de datos, allí se debe definir que data existe, las capacidades tecnológicas del negocio y la clasificación de la información, luego encontramos las operaciones fundamentales que permiten transacciones y análisis básicos, limpieza de datos, clasificación de datos validos o erróneos, entre otros tratamientos de información. Luego viene en esta pirámide la inteligencia de negocios donde se hace un análisis de la organización, se define los puntos clave del negocio proporcionarán esta información y sobre qué áreas clave del negocio se verán reflejados los beneficios de la implementación. Finalmente, y en la parte superior encontramos el Machine Learning que representa el nivel superior, es allí donde se implementará los modelos necesarios para analizar y procesar esta data de acuerdo a los objetivos planteados por la empresa. La tecnología de Machine Learning depende de una base sólida de datos y resalta la importancia de centrarse en fundamentos de datos antes de llegar a invertir en tecnología o sistemas de información, en conclusión, siempre es necesario e importante partir desde la base de dicha pirámide. [3]



Ilustración 1 Pirámide de Valor - Origen Crehana Introducción al Machine Learning

Luego se tiene la exposición de que el mayor desafío para las empresas es lograr que el 87% de los proyectos de Machine Learning puedan ser implementados, en muchos casos, este proceso se ve afectado por el origen de los datos, ya que en los ámbitos empresariales los datos son privados, o en algunas organizaciones no se cuenta con datos históricos de ninguna índole, en otros casos se tiene que los datos son solo fotografías,

audios, videos, u otros archivos multimedia. De allí surge la idea de que existen varios tipos de datos, como lo son los datos estructurados y los no estructurados.

El curso destaca la complejidad de los procesos y técnicas de procesamiento de datos, resaltando que siempre es necesario el manejo adecuado de los datos, verificación de los mismos, entre otros. En esta exposición, se menciona el concepto de la "pirámide de valor de los datos" como la mejor estrategia para implementar un proyecto con bases sólidas, avanzando en su implementación y siempre buscando generar mayor valor para la organización. [4]

Dentro del análisis de la exposición, se propone que existe un desafío para el 30% de las empresas con iniciativas en machine learning, por un lado, la escasez de talento, y luego de esto por la gestión efectiva de los datos. En muchos casos por temas educativos, en otros por la dificultad que representa el análisis de datos, o sencillamente porque los proyectos no fueron planteados con bases sólidas donde se garantice un beneficio o un valor agregado para la organización. [5]

Dentro de las recomendaciones principales es importante tener en cuenta que el machine learning se enfoca en el aprendizaje de acuerdo al análisis de los diferentes registros o iteraciones que ha tenido el usuario con el sistema previamente, y tomando como base los datos ya recolectados. Para garantizar una funcionalidad, es necesario definir una variable objetivo, dicha variable será un dato que se pueda "predecir" o estimar una probabilidad de éxito cuando se cumplen ciertas combinaciones de variables, o definir una variable objetivo que clasifique alguna acción como válida o inválida dependiendo de la naturaleza del negocio. De esta forma funcionan la mayoría de aplicaciones de machine learning con los datos recolectados de la experiencia de muchas personas o de muchos datos almacenados de iteraciones pasadas. [6]

La mayoría del machine learning funciona con tablas de datos con una variable objetivo y varios datos de variables definidas tales como fechas, valores, resúmenes, cálculos de datos, presupuestos, etc. Es importante tener en claro la variable objetivo o variable de interés que será calculada, o será predicha de acuerdo a las iteraciones de los diferentes usuarios, y entre mayor cantidad de iteraciones haya previamente almacenadas, los resultados serán de cierta forma más acertados. Dependiendo del objetivo, los tipos de variables y el tipo de problema, es aquí donde nos encontramos con un planteamiento diferente, y vamos a tener una diferencia entre los diferentes tipos de modelos, ya que podremos realizar ejercicios de clasificación o ejercicios de regresión.

Los ejercicios de clasificación se enfocan en definir un valor binario dependiendo de las variables o dependiendo de la necesidad del negocio, un ejercicio de clasificación identificará si es verdadero o falso, positivo o negativo, 1 o 0, aumenta o disminuye, es texto o número, etc. Básicamente tendrá un resultado ya preestablecido o definido en una lista de opciones.

Los ejercicios de regresión están enfocados en mostrar un comportamiento estimado de acuerdo al análisis de datos, es decir, el resultado obtenido puede variar dependiendo de

los cálculos dependiendo de los rangos, etc. Por ejemplo, un ejercicio de regresión puede estimar la temperatura de un día, lo que significa que el valor puede estar dentro de un rango estimado de temperaturas de acuerdo a las diferentes variables o registros previos, lo que confirma que no es una decisión binaria como los ejercicios de clasificación. Para los ejercicios de regresión, la técnica de regresión lineal es una forma básica para estimar un dato de acuerdo a los datos ya almacenados, esta función siempre estará buscando hallar el valor más cercano a todos los datos proporcionados o calculados dentro de la base de datos. [7] [8]

Es importante tener en cuenta que existen diferentes tipos de datos, como lo son los datos estructurados y no estructurados.

Los datos estructurados son datos que fácilmente se pueden tabular en forma de filas y columnas, donde en cada columna, de principio a fin, encontraremos un mismo tipo de dato, ejemplo una fecha, un valor, un promedio, un porcentaje, esto significa que para su análisis se puede implementar métodos matemáticos o estadísticos.

Los datos no estructurados son archivos con información más compleja de analizar, revisar y procesar ya que por su naturaleza, no es fácil encontrar patrones numéricos o secuencias lógicas. Por ejemplo, en el uso de datos no estructurados tenemos el procesamiento de audios, videos, imágenes u otros tipos de archivos multimedia.

Este tipo de procesamiento de datos no estructurados, para los humanos ha sido fácil desde hace mucho tiempo, pero para las maquinas no, hasta hace pocos años las maquinas han podido “aprender” a analizar este tipo de datos, por otro lado, desde hace varias décadas, las maquinas son mucho más ágiles que los humanos para realizar cálculos matemáticos, graficar datos estructurados, incluso, predecir datos de acuerdo a todo su procesamiento.

En resumen, el machine learning busca facilitar los procesos de entrenamiento, y aumentar la eficiencia de los sistemas de predicción, cálculos, clasificaciones, siempre enfocados en que haya una mayor confiabilidad de la información proporcionada por el sistema. [9]

2. Fundamentos aplicados de Machine Learning

En estos fundamentos o en esta etapa se procederá a instalar las herramientas necesarias para los ejercicios básicos del machine learning, para esto se requiere conocimientos base del lenguaje Python, se usará con las librerías numpy para la computación de los datos, usaremos pandas para el procesamiento de datos, usaremos scikits learn para el análisis predictivo y Matplotlib para la visualización grafica de los datos.

Como entorno de desarrollo usaremos Jupyter notebooks para las ecuaciones, ya que esta aplicación nos permite mezclar texto y código en un único archivo, facilitando la documentación. [10]

Como resumen del módulo para estimación del valor de una vivienda, se importó base de datos de Boston, se depuró la información, se creó una variable objetivo, se realizó un análisis de regresión y un modelo de regresión para estimar la relación entre las diferentes variables y el precio de venta de una casa.

Se parametrizo el entrenamiento del modelo con una parte de la base de datos y se hizo pruebas de valores de forma aleatoria para estimar el cálculo del precio de la vivienda. [11]

Como resumen del módulo aprobación de tarjetas de crédito, se realiza importación de base de datos local, se verifica toda la información y las inconsistencias, se revisa valores incompletos o nulos, se reemplazan por valores aceptables para la base de datos.

Luego de esto, se realiza las configuraciones para realizar el ejercicio de clasificación. Se define el modelo de clasificación, se segmenta la base de datos entre valores para el aprendizaje de modelo, así como la cantidad de valores que serían evaluados por el modelo basado en los datos de entrenamiento. [12]

El ciclo de vida del proyecto

En esta parte se debe analizar el proyecto y el valor que se desea agregar con el proyecto de implementación, en esta parte del módulo se verá el panorama más amplio aparte de los ejercicios de clasificación y regresión.

Para este análisis se debe conocer todo el negocio, los Kpi de negocio, los indicadores o los datos que podemos recolectar de las operaciones de las empresas.

Se debe revisar los objetivos definidos, y luego recolectar datos, en algunos negocios ya se encuentra la data, pero en otros, tan solo se debe empezar de cero con la data que se necesita para la implementación del proyecto.

Una vez obtenida la data, se debe definir el modelo, definir los parámetros de entrenamiento del algoritmo buscando la mejor predicción posible. En la mayoría de proyectos los algoritmos pueden ser mucho más complejos que los utilizados en la regresión y clasificación del curso.

Entrenar el modelo es la mitad del proceso, pero adicional es necesario que el modelo pase al proceso productivo, realizando predicciones de forma continua en ambiente real, cumpliendo el proceso completo de implementación, desarrollo, apis, plataformas donde será usado, etc. En conclusión, el desarrollo del modelo, es solo una pequeña parte del proyecto en general.

Existen otros tipos de algoritmos y técnicas de machine learning, no siempre los datos son datos estructurados, estos ejercicios con datos estructurados facilitan el modelo de pronostico y fortalecen el aprendizaje supervisado, pues los datos proporcionados para el entrenamiento son datos reales de la organización. [13]

Por otro lado, existen técnicas de aprendizajes no supervisados, donde no existen variable asociadas, por ejemplo, se puede agrupar datos por características similares, esto permite identificar las inconsistencias en los datos, o permiten la clasificación y agrupación de clientes, generando así una segmentación de los datos y apoyen el objetivo del negocio, todas estas segmentaciones deben estar aliadas con el conocimiento del tipo de negocio y objetivos. [14]

Existen otros tipos de algoritmos disponibles dentro del machine learning, tenemos por ejemplo las redes neuronales, arboles de decisiones, Support Vector Machine, aprendizaje Bayesiano, entre otros tipos de algoritmos. Las redes neuronales se definen como redes entre variables y el mapeo que se realiza para poder realizar la predicción al final, es muy utilizado en modelos supervisados, las redes realizan agrupaciones capa tras capa, agrupando datos que podrán influir dentro de la variable objetivo. [15]

En conclusión, a mayor volumen de información o datos, se garantiza un mejor entrenamiento y una mayor cercanía a la predicción planeada, sin embargo, el enfoque siempre será en resolver el problema objetivo con la técnica más adecuada.

3. Introducción a la ética en la inteligencia artificial

En los últimos años con el auge de la inteligencia artificial, se ha generado debates sobre la ética que debe tener principios éticos. Todo tipo de nueva tecnología generará estos debates para definir como se usará, beneficiará o afectará a toda la humanidad.

La forma en como los modelos han sido entrenados, es a través del conjunto de datos con que se entrenó, así que depende de estos datos las aplicaciones de sesgos o no en un modelo. Otra preocupación es la violación de los derechos de autor de todos los tipos de artes o profesiones, ya que varias inteligencias artificiales pueden tomar obras de arte, canciones, imágenes y hacer posibles violaciones a los derechos del autor.

Muchos modelos han sido entrenados con sesgos directos o indirectos, pero si un modelo ha sido entrenado con solo una parte de la información y no con todas las variables posibles, puede generar discriminaciones, sesgos, o incluso violaciones de los derechos de autor. [16]

El tema de exigir la publicación de las estructuras o programación de los algoritmos, no representa la solución a las infracciones éticas de las inteligencias artificiales. A nivel gubernamental internacional, es importante definir mecanismos de regulación y mecanismos de auditoria, control, y supervisión de los principios éticos. Esta auditoría de los principios éticos se debe hacer de forma articulada buscando prevenir la creación de herramientas de alto riesgo a nivel global, adicionalmente depende de nosotros como

población, también aprender a reconocer y filtrar aquellos sistemas que pueden vulnerar principios éticos o generar afectación a los demás miembros de la sociedad. [17]

Google cuenta con 7 principios sobre la aplicación de la inteligencia artificial, entre ellos tenemos

- Ser socialmente beneficioso, evitar crear o reforzar prejuicios injustos y responsabilizarse ante las personas afectadas. [18]

Como conclusión del curso, es bueno tener presente que, de nosotros como humanos pensantes, depende el uso que le demos a las herramientas creadas, evitando afectar a todos quienes nos rodean, incluyéndonos a nosotros mismos.

4. Evaluación de modelos de Machine Learning

Iniciando este curso se evidencia los modelos y los diferentes elementos básicos del machine learning.

Inicialmente tenemos las Etapas del workflow de un modelo de datos

- Entendimiento del negocio
- Entendimiento de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Puesta en producción [19]

Etapas de entrenamiento de un modelo de machine learning.

Para esto usaremos google colab que nos permitirá trabajar con los notebooks de forma sencilla y en línea, para esto usaremos un dataset descargado.

La preparación del dataset es lo más importante antes de empezar a entrenar los modelos, ya que con esta información es que se harán los cálculos para los pronósticos o clasificaciones que hará el modelo. [20]

La etapa de entrenamiento de un modelo consumirá un 20% de todo el tiempo del proyecto, ya que en este entrenamiento es donde se prepara el modelo para su funcionamiento futuro, esta es la parte más interesante del proyecto. Para esto tomaremos una parte del dataset para entrenamiento y se harán las pruebas para validar nuestro modelo. Como recomendación, no se debe utilizar en los datos de prueba los mismos datos con los que fue entrenado el modelo. [21]

Vamos a revisar las métricas de evaluación para un modelo de clasificación usadas por los científicos de datos.

Matriz de confusión:

Es una técnica para resumir el desempeño de un algoritmo de clasificación y los tipos de errores que está cometiendo, se trata de hacer una evaluación con un dataset con los valores esperados y se hace el análisis de los aciertos o desaciertos de nuestro modelo predictivo. A continuación, los valores que se evalúan en la matriz de confusión.

Dentro de esta matriz, las filas se entienden como las observaciones verdaderas, las predicciones se relacionan en las columnas. [22]

		PREDICCIÓN	
		Positivos	Negativos
OBSERVACIÓN	Positivos	Verdaderos positivos (VP)	Falsos negativos (FN)
	Negativos	Falsos positivos (FP)	Verdaderos negativos (VN)

Ilustración 2 Matriz de Confusión – Origen Crehana Evaluación de modelos de Machine

Learning

Accuracy:

Se deriva de la matriz de confusión y nos ayuda a tener una medida objetiva del desempeño de nuestro modelo, es una métrica que resume el modelo de clasificación como el número de predicciones correctas dividido en número total de predicciones, este resultado se puede ver como un porcentaje o como un fraccionario.

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

Ilustración 3 – Ecuación Accuracy Origen Crehana Evaluación de modelos de Machine Learning

La principal desventaja es que enmascara el problema de desequilibrio de clases, por ejemplo, si los datos contienen solo el 10% de las instancias positivas, un clasificador que asigne siempre la etiqueta negativa, asignaría la clase negativa al 90% del resto de las instancias, esto podría mostrar unas métricas poco acertadas. [23]

Precisión y Recall:

Esta métrica soluciona la desventaja del Accuracy, esta métrica presume el rendimiento del modelo de clasificación cuando se tienen dos clases con tamaño desigual, se calcula con el número de verdaderos positivos dividido en el número total de verdaderos positivos y falsos positivos.

También definimos el Recall como el número de verdaderos positivos dividido en el número total de verdaderos positivos y falsos negativos.

Esto mide que tan eficiente es nuestro modelo con la clasificación de los casos positivos. [24]

Precisión = $\frac{TP}{TP + FP}$

Recall = $\frac{TP}{TP + FN}$

F1 = 2 $\frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$

TP = True positive
TN = True negative
FP = False positive
FN = False negative

Ilustración 4 – Ecuación Precisión Recall y F1 Origen Crehana Evaluación de modelos de Machine Learning

Specificity y f1 Score:

Estas dos métricas son usadas en modelos de clasificación, y son métricas que se pueden calcular a partir de la matriz de confusión,

Specificity = $\frac{TN}{TN + FP}$

Ilustración 5 – Ecuación Specificity Origen Crehana Evaluación de modelos de Machine Learning

Para la métrica F1 Score es la medida de precisión y recall, unificando las dos métricas en una sola con la siguiente operación aritmética. F1 score tiene en cuenta tanto la precisión como el recall. [25]

F₁ = $2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

Ilustración 6 - Ecuación F1 Origen Crehana Evaluación de modelos de Machine Learning

Curva ROC y AUC

La curva roc es una medida de rendimiento para los problemas de clasificación en varias configuraciones del umbral de probabilidad, esta métrica es una curva de probabilidad y el área bajo la curva representa el grado o medida de separabilidad, es decir, nos indica cuanto es capaz el modelo de distinguir entre clases, cuanto sea mayor el área de la curva, mejor será el modelo para predecir la clase cero como clase cero, y la clase uno como clase uno.

La curva roc se caracteriza por el área bajo la curva, es decir a mayor área bajo la curva, mejor desempeño tendrá nuestro modelo. El área bajo la curva se puede tomar como un resumen del rendimiento del modelo. [26]

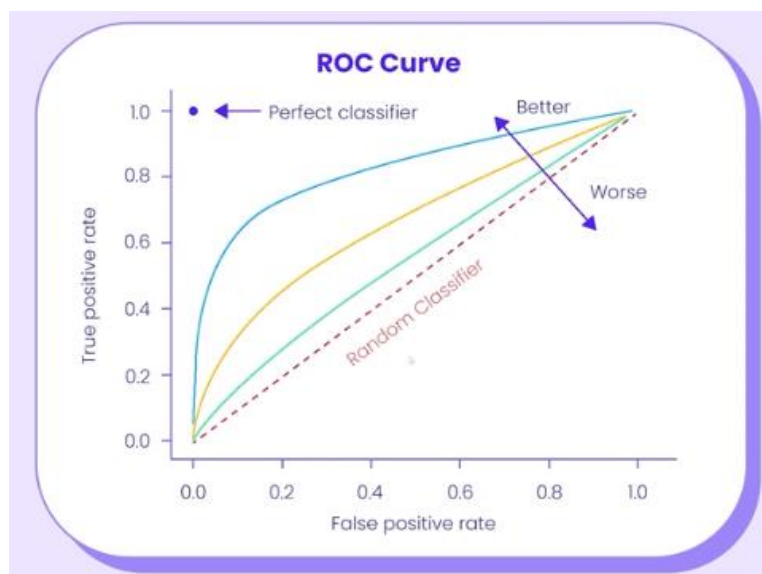


Ilustración 7 – Curva ROC Origen Crehana Evaluación de modelos de Machine Learning

Curva de precisión – Recall

La curva de precisión recall se utiliza para evaluar el rendimiento de los algoritmos de clasificación binaria, la curva precisión recall brinda una representación gráfica del rendimiento de un clasificador a través de muchos umbrales en lugar de un solo valor. Esta curva se construye calculando y graficando precisión contra recall para un solo clasificador en una variedad de umbrales. Por ejemplo, si usamos la regresión logística el umbral sería la probabilidad, que con observación pertenezca a la clase positiva. [27]

Existen técnicas de validación cruzada, donde se busca dar estabilidad a los errores del modelo, cuando un modelo presenta sesgos tiene un valor alto en todas las muestras, tanto en entrenamiento, validación y test.

Lo óptimo sería un modelo con poco sesgo y poca varianza, esto define el modelo predictivo como un modelo robusto y un buen balance.

Existen métodos para evaluar que tan bien clasificado están los datos por nuestro modelo, la validación cruzada es un método de remuestreo que utiliza diferentes partes de los datos para probar y entrenar un modelo en diferentes iteraciones, se utiliza en entornos donde el objetivo es la predicción, y se quiere estimar con que precisión se desempeñará un modelo predictivo en ejecución. [28]

5. Introducción a la Inteligencia Artificial

Las inteligencias artificiales en su auge intentan superar sus propias capacidades, sin embargo, tienen sus propias limitantes, tenemos la inteligencia artificial general que intenta imitar las capacidades cognitivas que tenemos los seres humanos y tenemos la inteligencia artificial narrow, que se enfoca en hacer actividades específicas de los seres humanos, actividades como reconocer texto, interpretar audios, reconocer imágenes o clasificar, incluso intuir o pronosticar algo. [29]

La inteligencia artificial AI Es la capacidad que tienen las máquinas para interpretar las capacidades cognitivas de los seres humanos, como entender lenguaje natural, entender imágenes, textos, audios. El Machine Learning es la herramienta que permite enseñarle a las computadoras a interpretar estas características que llevamos a cabo como seres humanos.

A las computadoras se les enseña a través de muchos ejemplos y muchos datos y a través de la repetición hasta que el sistema sea capaz de procesar datos por sí mismo y clasificarlos de acuerdo a su entrenamiento.

Deep Learning: Es una capacidad matemática, donde a través de métodos matemáticos y estadísticos el sistema calcula, extrae y analiza los diferentes resultados. De acá nace la ciencia de datos de todo este procesamiento de información, y de acá también surge el llamado big data con datos estructurados, semiestructurados y datos no estructurados. Las técnicas de Deep learning han ayudado a que las máquinas puedan identificar, reconocer, describir e interactuar con los seres humanos, estas técnicas han mejorado la precisión de los modelos.

Todo este Deep learning aprovecha toda la data existente y proporcionada por los usuarios, esto requiere gran volumen de procesamiento, almacenamiento y cálculo. [30]

Redes Neuronales: Estos modelos computacionales se asemejan a las neuronas y su funcionamiento cuando se transmite información de una neurona otra, en estas estructuras existe una entrada de información, un almacenamiento, un procesamiento y una réplica de información al siguiente nodo dentro de la red, las redes neuronales están divididas en capas que se dividen generalmente en tres tipos como la capa de entrada, una o más capas ocultas y la capa de salida. [31]

Compuertas Lógicas: Las compuertas lógicas trabajan en base de las tablas de la verdad OR y AND, en ellas se evalúa los datos y se determina un resultado lógico siempre y cuando cumplan las condiciones de acuerdo a las tablas de verdad. [32]

La arquitectura de las redes neuronales hace referencia al número de capas sino también al volumen de neuronas, existen diferentes tipos de arquitecturas como la feed forward, perceptron, feed forward networks entre otros. En sí, todas las arquitecturas buscan poder optimizar los procesos de análisis de datos.

Para el procesamiento de imágenes existe una arquitectura llamadas redes neuronales convulsionales para el procesamiento de imágenes, esta arquitectura se divide en bloques para el procesamiento de los datos. [33]

Dimensionalidad de los datos

Siempre es bueno tener presente las dimensiones de los datos por un lado conocer los datos de entrada y de la misma forma los datos de salida, dentro de la dimensionalidad de los datos tenemos datos escalares que suelen ser solo números o datos con un solo valor, vectores, datos en dos dimensiones con filas y columnas, datos de 3 y 4 dimensiones llamados tensores, normalmente las imágenes a color tienen una dimensión más que las imágenes a blanco y negro ya que cada color tiene una dimensión más sobre los matices que pueda tener. [34]

Casos de uso en la vida real de AI.

- **Visión por computadora:** Uno de los usos principales puede ser el reconocimiento facial, el reconocimiento de los autos Tesla que reconocen lo que tienen a su alrededor para saber por dónde deben conducir, este tipo de tecnología aún está en evolución.
- **Sistemas de recomendación:** Estos sistemas se entrenan de acuerdo al comportamiento de grandes grupos de usuarios, y agrupa tus comportamientos dentro de estos grandes grupos. De esta forma te brinda recomendaciones de acuerdo a su entrenamiento.
Spotify es una app que almacena los comportamientos de los usuarios para recomendarte las canciones que te pueden gustar, acá se usan análisis de audio para identificar tus posibles gustos.
Netflix también es una app que, de acuerdo a los comportamientos de grandes grupos de usuarios, recomienda de acuerdo a dichos gustos, allí se hacen análisis de imágenes, análisis de video, etc.
- **Evolución del NLP:** Este tipo de inteligencias enfocadas en el lenguaje natural se basan en el análisis estadístico del uso de las palabras, buscando cual es la mejor forma de responder, acá el machine learning tiene vital importancia porque de esta forma se entrena al sistema.
- **Procesamiento del lenguaje natural:** El procesamiento del lenguaje se basa en como entrenamos el sistema para que pueda responder, una parte se basa en las

reglas gramaticales, sin embargo, por la diversidad lingüística que existe, aumenta la dificultad en el proceso de entrenamiento de un sistema. [35]

GPT-3: Generative Pre-trained Transformer 3

Este algoritmo lanzado en mayo del 2020 por Open IA, ha sido entrenado con más de 175 billones de parámetros y la información en la internet por 8 años.

Esta herramienta es capaz de responder de una forma natural, es capaz de programar en el lenguaje solicitado por el usuario. Esta herramienta es de gran utilidad, aunque puede ser usada mal intencionadamente para generar cadenas de texto con noticias falsas donde no podríamos distinguir si fue escrito por un humano o por la máquina. [36]

Github Copilot: Esta herramienta desarrollada por GitHub y OpenAI para ayudar y asistir a los usuarios de Visual Studio Code, facilitando la inserción de código a partir de lenguaje natural, fue entrenado con la información de internet.

Esta herramienta facilita la inserción del código por medio de instrucciones en lenguaje natural, sugiriendo herramientas, funciones, importaciones, etc. [37]

6. Machine Learning: Análisis contrafactual

Contrafactual es el resultado de una acción que no sucedió, la inferencia causal se define en términos de un crontrafactual. El crontrafactual nos permitirá medir el efecto de alguna acción X sobre una variable Y, comparando como se hubiera comportado una variable en ausencia de la otra. [38]

Correlación vs Causalidad

La correlación es la presencia o ausencia de cualquier relación lineal entre 2 variables, esta correlación se puede atribuir a coincidencias entre variables, no siempre las relaciones entre variables son lógicas o son causadas por la otra, lo que significa que la correlación no implica causalidad. [39]

Uno de los problemas que existen al analizar los datos entre una variable u otra es el sesgo o algunos otros factores que afectan el resultado. Normalmente ocurre por el sesgo por variables omitidas dentro del análisis de datos. Las variables omitidas pueden ser observables o no observables dentro de los datos. [40]

Data science es el uso de métodos científicos para la extracción de conocimiento y valor desde los datos, esto involucra los métodos para analizar los datos los métodos más comunes son la estadística, econometría y machine learning, se busca siempre garantizar el análisis y la relación que existen entre las variables analizadas, sin embargo, hay variables que no tienen una relación directa y puede que no se evidencien variables causales de comportamientos no analizados. [41]

Los métodos estadísticos para inferir relaciones causales en los datos, es la modalidad que se utiliza para analizar el comportamiento de una variable dependiendo de la inferencia de otra variable, para estos análisis de datos se implementan varios métodos matemáticos uno de los métodos más usados es la regresión lineal. También existen retos para evaluar esta inferencia causal, sin embargo, hay varios métodos que se pueden usar para esta evaluación.

Uno de los métodos para hacer el análisis de inferencia causal es el A/B testing vs experimentos aleatorizados, esta técnica permite probar con agilidad conceptos básicos antes de diseñar nuevas operaciones, nuevos productos, etc.

El A/B testing clasifica toda la población muestra en dos grupos, grupo de control y grupo de tratamiento, y los experimentos aleatorizados, pueden evaluar a la vez varios grupos de tratamiento, estos experimentos se garantizan haciendo clasificaciones estratificadas para evitar desbalance en los grupos, esto evitaría sesgos también en el análisis del resultado obtenido después del experimento.

El método del matching busca balancear los dos grupos, el grupo de tratamiento y grupo de control siendo muy claro con la asignación de los miembros de cada grupo para minimizar el sesgo por clasificación. [42]

Tenemos también los efectos causales apalancados con el big data y del machine learning buscando minimizar la afectación por sesgo en las variables omitidas y la dificultad de elaborar contrafactuales, existe una ventaja y con el big data se tendrá acceso a muchos datos y variables que en muchos casos no son tenidas en cuenta dentro del análisis. Existen algoritmos para minimizar estas afectaciones, entre estos tenemos el modelo Double LASSO, este permite seleccionar las variables que más explican la variable de interés y reduciendo el sesgo por variables omitidas. [43]

Propensity Score es una técnica que ayuda a estimar efectos causales, esta técnica ayuda a balancear toda la cantidad de variables, para evitar analizar todas las variables al mismo tiempo, de esta forma se podrá analizar los promedios por resultados individuales buscando un balance. [44]

También tenemos la técnica llamada Double LASSO, esta evita analizar todas las variables, adicionalmente permite seleccionar las variables de mayor interés para las métricas a analizar evitando tener sesgos en las variables, este método funciona para la selección de variables y la estimación de coeficientes en modelos de regresión. [44]

7. Machine Learning: Aprendizaje supervisado

El machine learning es la base de la inteligencia artificial, y se clasifica en supervisado, no supervisado y semisupervisado o de aprendizaje profundo. Los datos tienen una gran importancia en machine learning, ya que los datos son los que permitirán entrenar el algoritmo, y hacer las clasificaciones o predicciones que se esperan del modelo. Los datos se clasifican en datos estructurados (organizados en tablas con filas y columnas) y los datos no estructurados como fotografías, audios, videos, etc. [46]

Para el procesamiento de estos datos dentro del machine learning supervisado, es necesario tener diferentes métricas de desempeño, de tal forma que se garantice que el mejor algoritmo será el que tenga mejor desempeño de acuerdo a las métricas, para las variables objetivo continuas o de regresión tenemos métricas como MAE, MSE y RMSE y para las variables objetivo categorías o de clasificación tenemos métricas como ACCURACY, F1 y curvas de ROC. [47]

Uno de los métodos matemáticos más usados es la regresión lineal, lo que se busca es minimizar el error, y encontrar los puntos más cercanos dentro de un modelo de predicción, también tenemos la regresión logística para el procesamiento de los datos.

Los árboles de decisión, en el ámbito de la ciencia de datos y aprendizaje automático, se destacan como algoritmos fundamentales que facilitan la toma de decisiones a partir de conjuntos de datos complejos. Estos algoritmos trabajan mediante la creación de estructuras de árboles, donde cada nodo representa una condición basada en atributos del conjunto de datos. La segmentación resultante ayuda en la clasificación de datos, ya que cada rama del árbol se asocia con una decisión específica basada en criterios condicionales. Así, los árboles de decisión ofrecen una manera intuitiva y eficaz de interpretar y utilizar la información contenida en los datos para realizar tareas de clasificación. [48]

Un algoritmo de Random Forest es una técnica de aprendizaje automático que utiliza múltiples árboles de decisión generados de manera aleatoria. El ensamblaje de los árboles se realiza mediante una ponderación basada en la precisión de cada uno en el conjunto de entrenamiento, lo que resulta en un modelo más robusto y preciso, capaz de abordar diversos problemas de manera efectiva. Existen algoritmos Random Forest continuos y categóricos. [49]

Una red neuronal es un subcampo del machine learning semisupervisado, se puede aplicar para predecir por medio de regresión o clasificación, una red neuronal es un algoritmo que cumple un proceso de tomar información a través de una capa input, y que luego se procesará por otras capas antes de obtener una capa de salida o capa output, en las redes neuronales, tenemos ciertas funciones llamadas funciones de activación, las cuales van a permitir el aprendizaje dentro de la red, estas funciones son diferentes a los nodos.

La arquitectura de neural network de las redes neuronales, la arquitectura subyacente guía la estructura y el funcionamiento de estas potentes herramientas de aprendizaje automático. Las redes neuronales están organizadas en capas, donde las neuronas interactúan entre sí a través de funciones de activación. Estas capas, a menudo llamadas "layers", albergan nodos que procesan y transmiten información, permitiendo que la red aprenda patrones complejos. La diversidad en estas arquitecturas da lugar a diferentes tipos de redes neuronales, algunas más densas o profundas que otras, cada una adaptada a tareas específicas.

Una distinción fundamental en la categorización de estas redes es la clasificación en "neural networks continuas" diseñadas para manejar datos que cambian de manera suave y continuo, son ideales para variables numéricas como la temperatura o la velocidad, es decir machine learning de regresión. Y "neural networks categóricas", que son óptimas para datos que se dividen en categorías distintas, como colores o tipos de productos o machine learning de clasificación. Esta diferenciación resalta la capacidad de las redes neuronales para adaptarse a la naturaleza específica de los datos, permitiendo un procesamiento eficiente y preciso en función de las características particulares de la tarea en cuestión. [50]

8. Innovación tecnológica con inteligencia artificial

Con el creciente volumen de información y herramientas tecnológicas, surge la inteligencia artificial, y con ella muchas ideas de innovación que a su vez son la combinación de saberes, ciencia y creatividad. La inteligencia artificial es la búsqueda de que las máquinas puedan ejecutar parte de la capacidad humana. La creatividad puede ser apoyada con la división de una idea en pequeños pasos para poder lograr ese objetivo. [51]

Hay diferentes motivos por los que se quiere mejorar e innovar en cualquier ámbito profesional o laboral, es bueno plantear siempre un objetivo que sea alcanzable, normalmente las ideas innovadoras se plantean para mejorar los estados de las empresas, todo inicia con la pregunta de valor, esa pregunta es la que apuntará para alcanzar un objetivo y su vez, al responderla brindará una solución a la necesidad de la compañía. [52]

Siempre es bueno seleccionar los tipos de gráficos para la creación de dashboard en la exposición de datos, una herramienta para esto se llama matplotlib y shiny
Explicabilidad: Es importante poder explicar cómo los modelos llegan al resultado, para esto existen varias técnicas tales como permutation importance y shap values
En conclusión, siempre es importante tener un balance entre combinar los conocimientos técnicos y ser creativos para implementar soluciones. [53]

Desarrollo e implementación del aprendizaje

El objetivo de este proyecto es plantear una idea de machine learning que ofrezca solución a una problemática dentro de una organización enfocada al procesamiento de materias primas para la venta en un restaurante.

Planteamiento del problema:

La empresa abastece la bodega principal y hace las compras dependiendo del volumen de ventas, dependiendo de las festividades y dependiendo del día de la semana, para dicho planteamiento solo tenemos el dataset con el registro de todas las compras realizadas durante el año 2022 y 2023, se espera que, con el desarrollo del modelo, se pueda pronosticar la cantidad a comprar por cada producto para las compras del año 2024.

Se plantea la siguiente pregunta de valor:

¿En qué momento se debe montar un pedido de compra para productos de un almacén?

Para dar solución a esta pregunta y analizando el planteamiento del problema, nos encontramos con un ejercicio de regresión compuesto con clasificación, de regresión por que las cantidades a pedir serán variables dependientes de las fechas y del día de la semana, y de clasificación por que hay que detectar cuando un producto aplica o no para montar orden de compra.

Para esto es necesario revisar la data existente y se cuenta con el registro de las compras de todo el año 2022 y 2023, en dicha data se tiene la cantidad comprada para cada día del año de los diferentes productos almacenados.

Para iniciar tendremos que definir los principios éticos que tendrá nuestro modelo de machine learning.

Principios Éticos:

1. El modelo no alterará los datos existentes previos.
2. El modelo no extraerá información no autorizada más que la proporcionada en el dataset.
3. La información que será proporcionada será de uso interno para la compañía de forma confidencial.
4. El desarrollo del modelo, estará enfocado solamente en abastecer la bodega principal del negocio.

El dataset disponible "Compras por día 2022 y 2023.csv" para hacer la predicción de compra por ítem consta de 40600 registros, 10 variables y solo dos variables de interés como lo es la Fecha de compra y la Cantidad en Entradas. A continuación, presento una imagen del dataset disponible.


```

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

# Cargar datos
data = pd.read_csv("Compras por día 2022 y 2023.csv")

# Filtrar datos para 2022 y 2023
data['Fecha'] = pd.to_datetime(data['Fecha'])
data_2022_2023 = data[(data['Fecha'].dt.year == 2022) | (data['Fecha'].dt.year == 2023)]

# Variables de interés
X = data_2022_2023[['Fecha', 'Cantidad Comprada']]
y = data_2022_2023['Aplica compra?']

# Crear conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Escalar los datos
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Modelo de regresión logística
model_logistic = LogisticRegression()
model_logistic.fit(X_train_scaled, y_train)
y_pred_logistic = model_logistic.predict(X_test_scaled)

# Matriz de confusión y métricas de regresión logística
conf_matrix_logistic = confusion_matrix(y_test, y_pred_logistic)
accuracy_logistic = accuracy_score(y_test, y_pred_logistic)
classification_report_logistic = classification_report(y_test, y_pred_logistic)

# Curva ROC de regresión logística
y_probs_logistic = model_logistic.predict_proba(X_test_scaled)[:, 1]
fpr_logistic, tpr_logistic, _ = roc_curve(y_test, y_probs_logistic)
roc_auc_logistic = auc(fpr_logistic, tpr_logistic)

# Modelo de red neuronal básica
model_nn = Sequential([
    Dense(64, activation='relu', input_dim=X_train_scaled.shape[1]),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid')
])
model_nn.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

```



```

model_nn.fit(X_train_scaled, y_train, epochs=10, batch_size=32, validation_split=0.2)
y_probs_nn = model_nn.predict(X_test_scaled)
y_pred_nn = (y_probs_nn > 0.5).astype("int32")

```

Matriz de confusión y métricas de red neuronal

```

conf_matrix_nn = confusion_matrix(y_test, y_pred_nn)
accuracy_nn = accuracy_score(y_test, y_pred_nn)
classification_report_nn = classification_report(y_test, y_pred_nn)

```

Curva ROC de red neuronal

```

fpr_nn, tpr_nn, _ = roc_curve(y_test, y_probs_nn)
roc_auc_nn = auc(fpr_nn, tpr_nn)

```

Graficar variación de compra por mes

```

data_2022_2023['Mes'] = data_2022_2023['Fecha'].dt.month
monthly_variation = data_2022_2023.groupby(['Mes', 'Aplica
compra?']).size().unstack().reset_index()
monthly_variation.plot(x='Mes', kind='bar', stacked=True, title='Variación de Compra
por Mes')
plt.xlabel('Mes')
plt.ylabel('Cantidad')
plt.show()

```

Graficar curvas ROC

```

plt.figure(figsize=(8, 6))
plt.plot(fpr_logistic, tpr_logistic, color='darkorange', lw=2, label=f'Curva ROC Regresión
Logística (AUC = {roc_auc_logistic:.2f})')
plt.plot(fpr_nn, tpr_nn, color='green', lw=2, label=f'Curva ROC Red Neuronal (AUC =
{roc_auc_nn:.2f})')
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlabel('Tasa de Falsos Positivos (FPR)')
plt.ylabel('Tasa de Verdaderos Positivos (TPR)')
plt.title('Curvas ROC')
plt.legend(loc='lower right')
plt.show()

```

Mostrar resultados y métricas

```

print("Matriz de Confusión - Regresión Logística:")
print(conf_matrix_logistic)
print("\nMétricas - Regresión Logística:")
print(f'Accuracy: {accuracy_logistic}')
print(f'Classification Report:\n{classification_report_logistic}\n')
print("Matriz de Confusión - Red Neuronal:")
print(conf_matrix_nn)

```

```
print("\nMétricas - Red Neuronal:")
print(f'Accuracy: {accuracy_nn}')
print(f'Classification Report:\n{classification_report_nn}')
```

Código Python para modelo de regresión:

Desarrollar en Python modelo predictivo de machine learning para *identificar cantidad a comprar de un producto en una bodega*.

- Datos disponibles: Dataset llamado "Compras por día 2022 y 2023.csv"
- Variables de Interés: Fecha y Cantidad Comprada
- Variable Objetivo: Cantidad a Comprar

Código Python generado por OpenAI ChatGPT 3.5 [54]

Importar librerías

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Cargar datos

```
data = pd.read_csv("Compras por día 2022 y 2023.csv")
```

Filtrar datos para 2022 y 2023

```
data['Fecha'] = pd.to_datetime(data['Fecha'])
data_2022_2023 = data[(data['Fecha'].dt.year == 2022) | (data['Fecha'].dt.year == 2023)]
```

Variables de interés

```
X = data_2022_2023[['Fecha']]
y = data_2022_2023['Cantidad Comprada']
```

Crear conjunto de entrenamiento y prueba

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Crear modelo de regresión lineal

```
model = LinearRegression()
```

Entrenar modelo

```
model.fit(X_train[['Fecha']], y_train)
```

Realizar pruebas de predicción

```
y_pred = model.predict(X_test[['Fecha']])
```

Evaluar el rendimiento del modelo

```
mse = mean_squared_error(y_test, y_pred)
```

```
r2 = r2_score(y_test, y_pred)
```

```
print(f'Mean Squared Error: {mse}')
```

```
print(f'R2 Score: {r2}')
```

Graficar variación de compra por cada mes

```
data_2022_2023['Mes'] = data_2022_2023['Fecha'].dt.month
```

```
monthly_variation = data_2022_2023.groupby('Mes')['Cantidad  
Comprada'].sum().reset_index()
```

```
plt.figure(figsize=(10, 6))
```

```
sns.barplot(x='Mes', y='Cantidad Comprada', data=monthly_variation)
```

```
plt.title('Variación de Compra por Mes')
```

```
plt.xlabel('Mes')
```

```
plt.ylabel('Cantidad Comprada')
```

```
plt.show()
```

Figuras y tablas

Ilustración 1 Pirámide de Valor - Origen Crehana Introducción al Machine Learning.....	6
Ilustración 2 Matriz de Confusión – Origen Crehana Evaluación de modelos de Machine Learning	12
Ilustración 3 – Ecuación Accuracy Origen Crehana Evaluación de modelos de Machine Learning	13
Ilustración 4 – Ecuación Precisión Recall y F1 Origen Crehana Evaluación de modelos de Machine Learning	14
Ilustración 5 – Ecuación Specificity Origen Crehana Evaluación de modelos de Machine Learning	14
Ilustración 6 - Ecuación F1 Origen Crehana Evaluación de modelos de Machine Learning	14
Ilustración 7 – Curva ROC Origen Crehana Evaluación de modelos de Machine Learning	15

Conclusiones

En la ejecución del seminario se pudo aprender, conocer e indagar sobre las tecnologías relacionadas con el machine learning (ML) y la inteligencia artificial (IA), es muy interesante entender como un sistema informático puede ser entrenado para predecir, calcular o clasificar una variable objetivo.

Siendo creativo, ético y de la mano de muchas herramientas de desarrollo como Python, matplotlib, pandas, seaborn, sklearn y muchas más, se puede desarrollar herramientas de gran utilidad para atender la necesidad en las empresas, se pudo conocer herramientas que generan valor para una compañía.

En el proyecto se aprendió y abordó desde los temas básicos e introductorios a la tecnología del machine learning, entendiendo como los algoritmos pueden “aprender” o ser entrenados con datos de iteraciones de muchos usuarios, pasando por ejercicios de aplicación del machine learning se adquieren bases sobre la construcción de modelos y entrenamiento de modelos a través de las diferentes herramientas como google colab y jupyter notebook disponibles al alcance de todos.

Se abordó la importancia que está tomando la inteligencia artificial en la carrera por replicar funciones o acciones humanas tales como procesar y entender imágenes, procesar y entender audios, videos o cualquier tipo de archivo multimedia, todo acompañado de modelos supervisados o no supervisados para garantizar el entrenamiento del sistema predictivo.

Dentro del seminario se pudo comprender la importancia de las métricas en el desarrollo de un aplicativo para garantizar confiabilidad, eficiencia y seguridad en los datos proporcionados. Es importante siempre aplicar estas técnicas que permitan evaluar los modelos desarrollados para garantizar calidad.

Se pudo comprender la importancia de ver el panorama completo incluyendo aquellas acciones que nunca tuvieron un resultado o acciones que nunca se materializaron como el análisis contrafactual, de cómo puede generar impacto un hecho que no ocurrió o como no tener en cuenta unas variables pueden afectar el resultado de un modelo computacional.

Fue posible reconocer la importancia de la ética dentro de nuestra profesión como desarrolladores y dentro de estas nuevas tecnologías de inteligencia artificial para no vulnerar derechos intelectuales de otros, o vulnerar la privacidad o tomar datos de fuentes no autorizadas, se aprendió sobre la importancia de ser ético en el momento de brindar, procesar o recibir información. Como reflexión, siempre es necesario garantizar que las tecnologías sean usadas en pro y beneficio de la humanidad.

Bibliografía

- [1] L. A. Lee, «Machine Learning en tu día a día,» de *Introducción a Machine Learning*, Agosto 2021.
- [2] L. A. Lee, «Entonces... ¿Qué es Machine Learning y por qué ahora?,» de *Introducción a Machine Learning*, Agosto 2021.
- [3] L. A. Lee, «Desmitificando...1,» de *Introducción a Machine Learning*, Agosto 2021.
- [4] L. A. Lee, «Desmitificando...2,» de *Introducción a Machine Learning*, Agosto 2021.
- [5] L. A. Lee, «Estrategia de Datos,» de *Introducción a Machine Learning*, Agosto 2021.
- [6] L. A. Lee, «Introducción a Machine Learning,» de *Introducción a Machine Learning*, Agosto 2021.
- [7] L. A. Lee, «Clasificación vs. Regresión,» de *Introducción a Machine Learning*, Agosto 2021.
- [8] L. A. Lee, «La básica y confiable, la regresión lineal,» de *Introducción a Machine Learning*, Agosto 2021.
- [9] L. A. Lee, «Los tipos de datos: ¿Tienen estructura?,» de *Introducción a Machine Learning*, Agosto 2021.
- [10] L. A. Lee, «Python, Paquetes y librerías,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [11] L. A. Lee, «Prediciendo el precio de una casa: Regresiones en Boston,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [12] L. A. Lee, «¿A quién le prestarías dinero? Clasificación para tarjetas de crédito,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [13] L. A. Lee, «El Ciclo de Vida del Proyecto,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [14] L. A. Lee, «¿Con o sin supervisión? Clustering y otros tipos de Machine Learning,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [15] L. A. Lee, «Deep Learning: Redes Neuronales,» de *Fundamentos aplicados de Machine Learning*, Octubre 2021.
- [16] E. Wohlmuth, «Principios éticos y desarrollo de algoritmos en modelos de inteligencia artificial,» de *Introducción a la ética en la Inteligencia Artificial*, Agosto 2023.
- [17] E. Wohlmuth, «Responsabilidad en el uso de diferentes modelos de IA,» de *Introducción a la ética en la Inteligencia Artificial*, Agosto 2023.
- [18] E. Wohlmuth, «Análisis de casos de éxito: Manejo responsable y transparente de la IA,» de *Introducción a la ética en la Inteligencia Artificial*, Agosto 2023.

- [19] M. Rojo, «Machine Learning Workflow,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [20] M. Rojo, «Preparación de datos (parte 1 y 2),» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [21] M. Rojo, «Modelado,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [22] M. Rojo, «Matriz de confusión,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [23] M. Rojo, «Accuracy,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [24] M. Rojo, «Precision y Recall,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [25] M. Rojo, «Specificity y F1 score,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [26] M. Rojo, «Curva ROC y AUC,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [27] M. Rojo, «Curva de Precision-Recall,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [28] M. Rojo, «Técnicas de validación cruzada,» de *Evaluación de modelos de Machine Learning*, Enero 2023.
- [29] F. D. R. Martínez, «Estado del arte de la inteligencia artificial,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [30] F. D. R. Martínez, «Diferencias entre AI, ML, DL y Data Science,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [31] F. D. R. Martínez, «Redes neuronales,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [32] F. D. R. Martínez, «Compuertas lógicas,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [33] F. D. R. Martínez, «Arquitectura de las redes neuronales,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [34] F. D. R. Martínez, «Dimensionalidad de los datos,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [35] F. D. R. Martínez, «Casos de uso en la vida real,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [36] F. D. R. Martínez, «GPT-3,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [37] F. D. R. Martínez, «Github Copilot,» de *Introducción a la Inteligencia Artificial*, Enero 2022.
- [38] C. Tabares, «¿Qué es un contrafactual?,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.

- [39] C. Tabares, «Correlación vs. causalidad,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [40] C. Tabares, «Variables omitidas y sesgo de selección,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [41] C. Tabares, «Causal Data Science,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [42] C. Tabares, «A/B testing vs. experimentos aleatorizados,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [43] C. Tabares, «Inferencia causal con Big Data,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [44] C. Tabares, «Propensity Score,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [45] C. Tabares, «Double LASSO,» de *Machine Learning: Análisis contrafactual*, Septiembre 2022.
- [46] H. A. Aragón, «Inteligencia artificial, Machine learning supervisado y no supervisado,» de *Machine Learning: Aprendizaje supervisado*, Octubre 2021.
- [47] H. A. Aragón, «Métricas de desempeño,» de *Machine Learning: Aprendizaje supervisado*, Octubre 2021.
- [48] H. A. Aragón, «Árboles de decisión - categóricos,» de *Machine Learning: Aprendizaje supervisado*, Octubre 2021.
- [49] H. A. Aragón, «Random forest,» de *Machine Learning: Aprendizaje supervisado*, Octubre 2021.
- [50] H. A. Aragón, «Introducción neural network,» de *Machine Learning: Aprendizaje supervisado*, Octubre 2021.
- [51] R. A. M. Garrido, «Beneficios de la IA en la innovación tecnológica,» de *Innovación tecnológica con inteligencia artificial*, Septiembre 2023.
- [52] R. A. M. Garrido, «Identificación de oportunidades,» de *Innovación tecnológica con inteligencia artificial*, Septiembre 2023.
- [53] R. A. M. Garrido, «Monitoreo y optimización continua con IA,» de *Innovación tecnológica con inteligencia artificial*, Septiembre 2023.
- [54] OpenAI, «Desarrollo de Modelo Predictivo Machine learning,» ChatGPT 3.5, 2023. [En línea].