

Universidad Nacional de General Sarmiento

LABORATORIO DE CONSTRUCCIÓN DE SOFTWARE

PROYECTO PROFESIONAL 1

TRABAJO PRÁCTICO INICIAL

ENTREGA 1

INTEGRANTES

Guadalupe Nicole Arroyo

Lautaro Manuel Avalos

Federico Emanuel Farias

PROFESORES

Ing. Juan Carlos Monteros

Ing. Francisco Orozco De La Hoz

Lic. Leandro Dikenstein

INTRODUCCIÓN

En un mundo cada vez más interconectado y complejo, la convergencia de la inteligencia artificial, el aprendizaje automático, el aprendizaje profundo y la ciencia de datos ha desencadenado avances sin precedentes en nuestra capacidad para comprender y abordar cuestiones trascendentales en la sociedad. Estas disciplinas, una vez consideradas tecnologías emergentes, han evolucionado para convertirse en herramientas fundamentales en la búsqueda de soluciones a problemas complejos.

A lo largo de este trabajo, explicaremos qué son y cómo funcionan estas disciplinas. Exploraremos cómo la conjunción de estas tecnologías permite analizar datos, refinarlos, utilizar y entrenar modelos de machine learning para clasificar o predecir eventos y sobre todo que objetivo o caso de uso usaremos para que sea útil para el equipo y la comunidad.

Utilizaremos el repositorio [LabSWPP12023S2G2](#) en Github para archivos de investigación, informes, datasets y código a futuro, [Anaconda Cloud](#) para prácticas con Jupyter Notebooks, Jira para las tareas del equipo y seguiremos con la investigación de más herramientas en el transcurso del trabajo práctico.

INTELIGENCIA ARTIFICIAL

La inteligencia artificial abarca múltiples disciplinas diferentes, como la informática, análisis de datos, estadísticas, neurociencia, ingeniería de hardware y software, entre otras. Podríamos definir a la inteligencia artificial como un conjunto de tecnologías implementadas en un sistema que puedan imitar comportamientos humanos diversos para resolver problemas complejos. Estas tecnologías se basan principalmente en el aprendizaje automático (Machine Learning) y el aprendizaje profundo (Deep Learning), usadas para el análisis de datos, la generación de predicciones y previsiones, la categorización de objetos, el procesamiento de lenguaje natural, las recomendaciones, la recuperación inteligente de datos y mucho más.

MACHINE LEARNING

Es una rama del campo de la IA capaz de reproducir un comportamiento utilizando algoritmos alimentados por una gran cantidad de datos. En lugar de una programación explícita, el aprendizaje automático usa algoritmos para analizar grandes cantidades de datos, aprender de las estadísticas y tomar decisiones fundamentadas. Los algoritmos de ML se pueden clasificar en los siguientes tipos:

Aprendizaje Supervisado:

Los algoritmos de aprendizaje supervisado utilizan ejemplos de entrada y salida etiquetados para aprender a realizar predicciones. Se dividen en 2 tipos:

- **Regresión:** Prevé valores numéricos continuos (como precios o cantidades) basándose en relaciones entre variables y algunos ejemplos son:
 - **Regresión Lineal:** Busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. El objetivo es predecir valores de la variable dependiente basados en los valores de las variables independientes. Se representa como una ecuación de una línea recta o un hiperplano.

- **Regresión Polinómica:** A veces, los datos no siguen una relación lineal. En este caso, la regresión polinómica se utiliza para modelar relaciones no lineales al introducir términos polinómicos en el modelo.
- **Regresión de Bosque Aleatorio:** Esta técnica combina múltiples árboles de decisión para obtener predicciones más robustas. Es útil cuando se quiere manejar la no linealidad y el ruido en los datos.
- **Clasificación:** Clasifica instancias en categorías predefinidas (como sí/no o clases) según patrones aprendidos en los datos de entrenamiento y algunos ejemplos son:
 - **Regresión Logística:** A pesar de su nombre, la regresión logística se utiliza para la clasificación. Calcula la probabilidad de que una instancia pertenezca a una clase específica y luego toma una decisión de clasificación basada en un umbral.
 - **Árboles de Decisión:** Dividen los datos en ramas en función de características para llegar a una decisión de clasificación en las hojas.
 - **Naïve Bayes:** Utiliza el teorema de Bayes para calcular la probabilidad de que una instancia pertenezca a una clase específica. Supone independencia entre las características, de ahí el término "naïve" (ingenuo).

Aprendizaje No Supervisado:

Los algoritmos de aprendizaje no supervisado descubren patrones en los datos sin utilizar etiquetas. Algunos ejemplos son:

- **Agrupamiento (Clustering):** Aquí, el objetivo es identificar patrones y similitudes en los datos sin etiquetas previas. Los algoritmos de agrupamiento agrupan instancias similares en clústeres, lo que puede ayudar a descubrir estructuras ocultas en los datos.
- **Reducción de Dimensionalidad:** Implica reducir la cantidad de variables o características en un conjunto de datos, manteniendo la mayor parte de la información relevante. Esto puede ayudar a simplificar los modelos y mejorar la eficiencia.
- **Asociación:** Estos algoritmos buscan encontrar relaciones y patrones interesantes entre diferentes elementos en un conjunto de datos. Son ampliamente utilizados en el análisis de la cesta de compras y la minería de datos.

Aprendizaje por Refuerzo:

El aprendizaje por refuerzo implica un agente que toma decisiones en un entorno para maximizar una recompensa acumulada. Un ejemplo es:

- **Refuerzo tabular:** En este tipo de enfoque, el espacio de estados y acciones del agente es manejable y se puede representar fácilmente en una tabla.
- **Refuerzo episódico y continuo:** El agente interactúa en episodios discretos, mientras que otros funcionan en entornos continuos donde la interacción no está dividida en episodios claramente definidos.

DEEP LEARNING

Es una rama del campo de la inteligencia artificial y el aprendizaje automático que se enfoca en el uso de redes neuronales artificiales profundas para aprender y extraer patrones complejos de datos. Estas redes están diseñadas para simular el funcionamiento del cerebro humano en la forma en que procesan y organizan la información. Ejemplos incluyen:

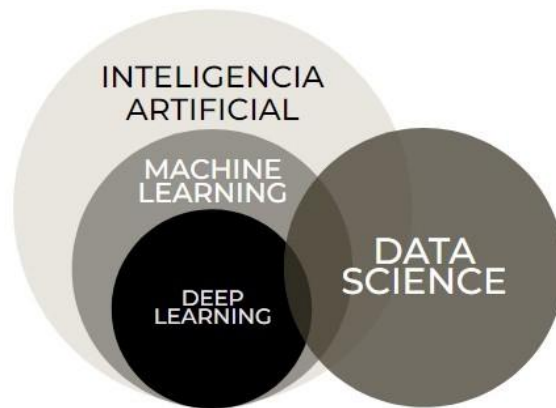
- **Redes Neuronales Convolucionales (CNN):** Utilizadas en tareas de visión por computadora, como el reconocimiento de objetos en imágenes.
- **Redes Neuronales Recurrentes (RNN):** Adecuadas para datos secuenciales, como el procesamiento de lenguaje natural y la generación de texto.

DATA SCIENCE

Es un campo multidisciplinario que implica procesar, analizar y extraer información para mejorar la toma de decisiones. Utiliza técnicas de análisis y visualización de datos, junto con modelos de IA, para encontrar patrones, predecir eventos y obtener conocimientos valiosos. Por ejemplo, un banco puede identificar clientes solventes, una empresa de viajes puede vincular vuelos y cruceros, y un laboratorio puede usar datos para desarrollar medicamentos eficaces contra virus.

CÓMO INTERACTÚAN ENTRE SÍ

Aquí podemos observar cómo cada disciplina es una subdisciplina de la anterior a excepción de Data Science, que utiliza a las anteriores como explicamos anteriormente.



OBJETIVO

Hemos elegido trabajar con el [dataset de la UNC de niveles y predictores de depresión, ansiedad y riesgo de suicidio durante la pandemia de COVID-19 en Argentina](#), el cual refinaremos provincia por provincia para clasificarlas según el potencial riesgo de suicidios de cada una. Luego, poder ayudar a la comunidad a crear programas específicos para mejorar aspectos de salud mental y prevención del suicidio en provincias que tengan mayor riesgo. Esto lo haremos mediante modelos de aprendizaje supervisado, como regresión logística o árboles de decisión.

CONCLUSIONES

Luego de la investigación inicial, nos encontramos con un gran abanico de posibilidades para abordar el trabajo y las futuras entregas. A causa de esto esperamos poder identificar y utilizar los modelos que mejor se adapten a nuestro dataset y poder cumplir el objetivo propuesto.

BIBLIOGRAFÍA

Wikipedia (<https://www.wikipedia.org/>)

Oracle (<https://www.oracle.com/>)

IBM (<http://www.ibm.com/>)

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

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

Dirección Nacional de Abordaje Integral de Salud Mental y Consumos Problemáticos. (s.f.), *Salud Mental*. Ministerio de Salud.

<https://www.argentina.gob.ar/salud/mental-y-adicciones/que-es>

Dirección Nacional de Abordaje Integral de Salud Mental y Consumos Problemáticos. (s.f.), *Suicidio - Prevención*. Ministerio de Salud.

<https://www.argentina.gob.ar/salud/mental-y-adicciones/suicidio>

López Steinmetz, L. C. (2021, May 3). *R Code and dataset for: Levels and predictors of depression, anxiety, and suicidal risk during COVID-19 pandemic in Argentina: The impacts of quarantine extensions on mental health state* [Dataset].

<https://rdu.unc.edu.ar/handle/11086/20168>