
Universidad
Tecmilenio®

Universidad Tecmilenio

Ética aplicada a la inteligencia artificial

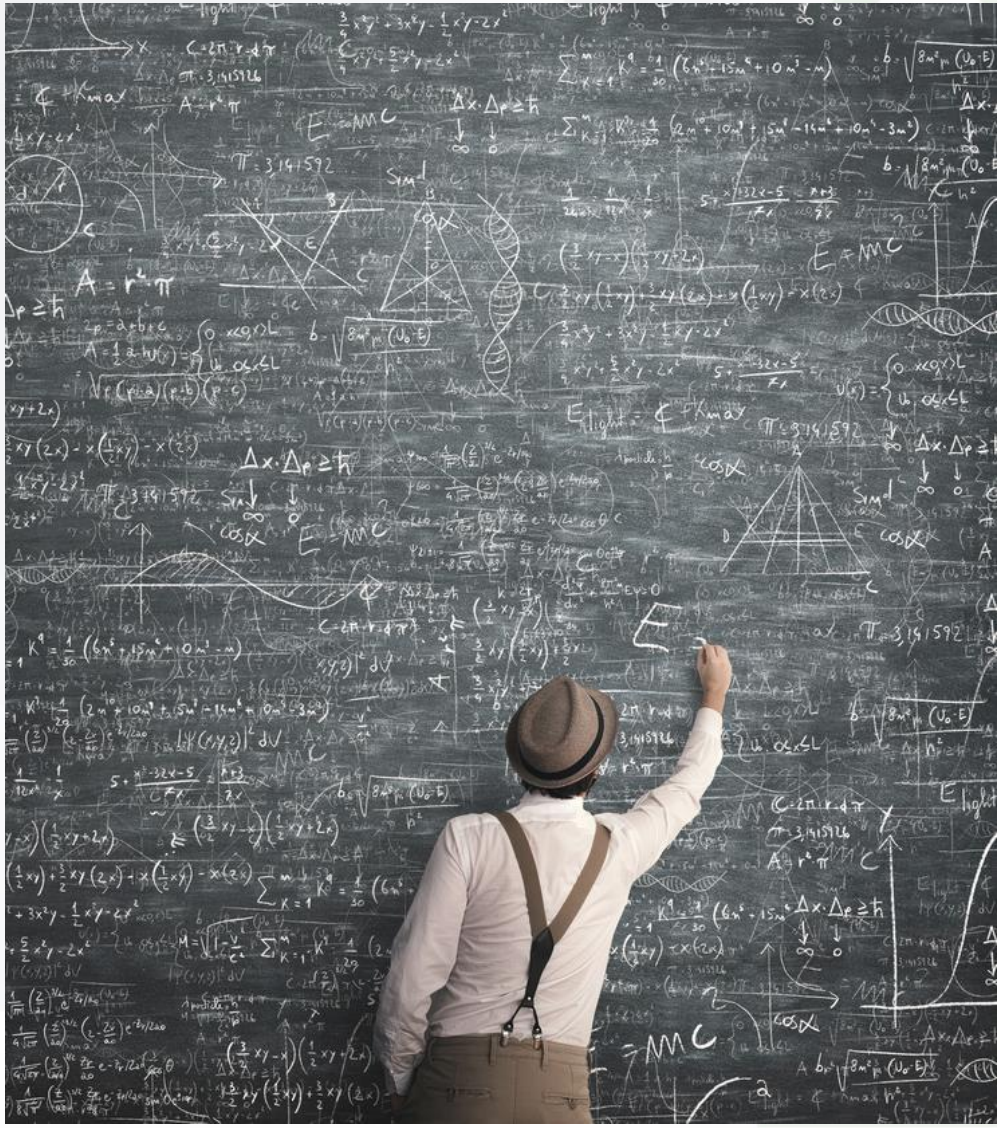
Evaluación de modelos

Introducción

Un ejemplo de cómo un algoritmo puede resolver una tarea de forma distinta, con un diferente nivel de complejidad, es el de ordenamiento. En realidad, si se consideraran 10 algoritmos de ordenamiento para construir uno nuevo, este tendría 10 maneras diferentes de resolver la misma tarea, y al final, se tendría que decidir con base en algún criterio cuál de ellos funcionó mejor para la lista de números de entrada que se le proveyó.

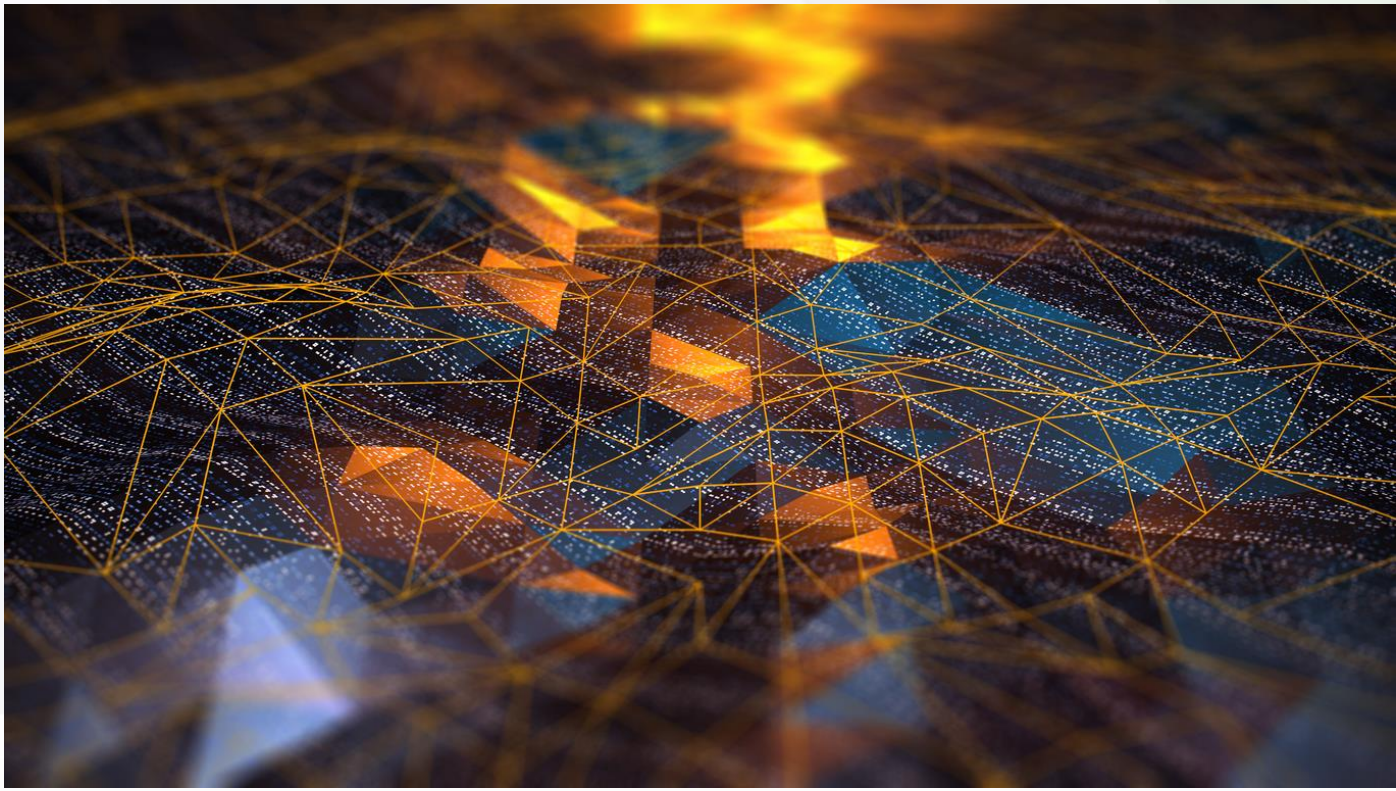
En este tema conocerás:

- Los modelos predictivos.
- El propósito de la precisión de un modelo.
- Las limitantes de los modelos predictivos.
- Los diferentes tipos de sesgo en los modelos.
- Los conceptos de desajuste y sobreajuste de modelos.



Explicación

La idea de que un algoritmo se construye a partir de sí mismo se conoce como **modelado**, el cual se puede ver como un algoritmo de aprendizaje que se deriva de los datos, en lugar de aprender (o realizar sus acciones) con instrucciones declaradas de forma explícita. La diferencia entre desarrollar un algoritmo y desarrollar un modelo es que para el primero solo se requiere escribir de forma explícita las instrucciones para la tarea que tiene que realizar, eventualmente se evalúa su desempeño y se modifican las instrucciones en caso de que algo no suceda de la forma esperada. Por otro lado, desarrollar un modelo implica escribir las instrucciones sobre la forma de modelar los datos, alimentarlos y con ello evaluar su desempeño, y en caso de algún resultado no deseado o incorrecto, ajustar el modelo. Al final, los algoritmos construyen modelos a partir de los datos.



Explicación

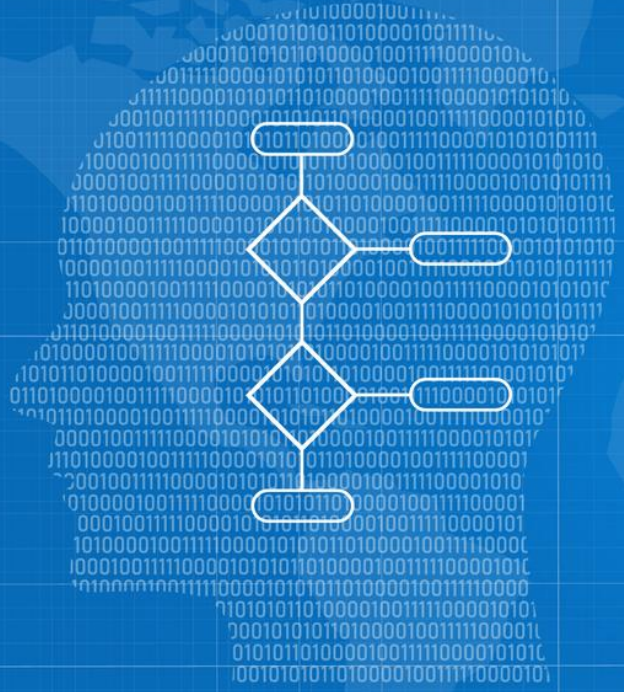
Al comparar las predicciones que hace el modelo sobre el mismo conjunto de datos de entrada, se obtiene una función de error, que no es otra cosa más que la comparación del valor de salida que el modelo predice sobre una instancia del conjunto de datos de entrada contra la etiqueta real de esa instancia. Si el error esta fuera de cierto criterio, entonces el modelo se ajusta con la finalidad de que en cada ciclo de predicción y comparación el error se minimice. Este proceso se conoce como **entrenamiento del modelo** y da como resultado el modelo que mejor se ajusta a los datos con los que fue entrenado.

Explicación

Una vez finalizada la etapa de entrenamiento, el modelo está listo para ser desplegado en el escenario real. Con la finalidad de evaluar qué tan bien puede generalizar (predecir) con datos nunca vistos, se utiliza un conjunto de datos distinto al que se utilizó en el entrenamiento. El resultado de esta evaluación permite conocer el índice de error del algoritmo, una forma de obtenerlo es con la **precisión.** La precisión indica qué tan bien el modelo es capaz de predecir la clase correcta en una instancia del conjunto de datos de entrada, y corresponde al número de instancias predichas correctamente entre todas las instancias del conjunto de datos.

Explicación

Una modelo con alta precisión no necesariamente significa que es estable o que aprendió con precisión. El **sesgo** en los modelos es un fenómeno que ocurre cuando un modelo produce resultados que tienen prejuicios sistémicos debido a supuestos erróneos durante su construcción (proceso de aprendizaje). Un modelo de aprendizaje automático depende de la calidad, objetividad y tamaño del conjunto de datos a partir del cual fue construido. Si el conjunto de datos es erróneo, pobre o está incompleto, el modelo generará predicciones erróneas. "*Garbage in, garbage out*" (Geiger et al., 2021))transmite la idea de que la calidad de la salida del sistema está determinada por la calidad de la entrada.



Algorithmic bias

Actividad

Describe tres ejemplos explicando de qué forma una regresión lineal puede ser útil para predecir algo en tu campo laboral o escolar. Puedes considerar el siguiente artículo sobre **regresión lineal:**

Cierre

Mientras los algoritmos de aprendizaje automático generan modelos que permiten que las compañías tengan procesos de toma de decisiones más eficientes y automatizadas, estos son susceptibles al síndrome de “garbage in, garbage out”, derivado de conjuntos de datos sesgados. El sesgo es un fenómeno que carga el resultado de un algoritmo o modelo a favor o en contra de una idea.

Para evitarlo, es importante seleccionar un conjunto de datos de entrenamiento que sea representativo y lo suficientemente grande para contrarrestar los tipos más comunes de sesgo, el muestral y el cognitivo. De igual manera, en la medida que pasa el tiempo, se deben actualizar los conjuntos de datos de entrenamiento y reentrenar los modelos para que aprendan los nuevos patrones de datos que pudieran aparecer.

Bibliografía

- Geiger, R., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., y Tang, R. (2021). “Garbage in, garbage out” revisited: What do machine learning application papers report about human-labeled training data? *Quantitative Science Studies*, *2*(2).

Universidad Tecmilenio

Ética aplicada a la inteligencia artificial

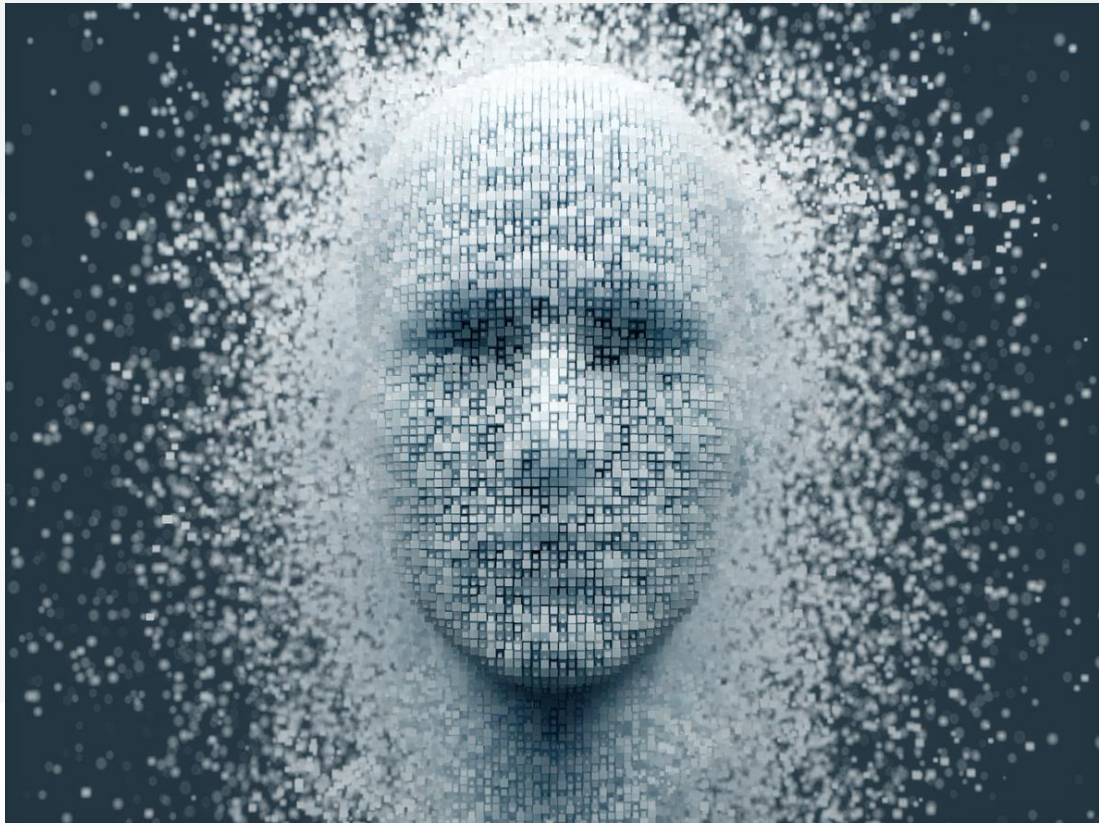
El derecho a la privacidad

Introducción

En el 2016 se crearon aproximadamente 44 billones de gigabytes de datos por día y se estima que para el 2025 se alcance la cifra de 463 billones de gigabytes diariamente. Más datos, más violaciones a la privacidad.
En la medida que más conjuntos de datos son vinculados a personas y compartidos en modelos de “código abierto”, se convierten en recursos sumamente atractivos.
En este tema conocerás:
La definición de privacidad en el contexto del aprendizaje automático.
Identificar problemas de privacidad en conjuntos de datos anónimos.
El concepto de anonimización.
Las diferentes técnicas para anonimizar un conjunto de datos.
La k-anonimidad.

Explicación

En 2012, el New York Times publicó una historia sobre la forma en que las empresas analizan la información recolectada de sus clientes con el fin de buscar ventajas publicitarias. El artículo reveló que Target Corporation en los Estados Unidos tenía un algoritmo capaz de determinar a partir de los patrones de compra si una mujer estaba embarazada y utilizar esa información para enviarle anuncios publicitarios y cupones de artículos relacionados con el bebé. El reporte incluía una anécdota sobre un hombre en Minneapolis que se había quejado porque la tienda le había enviado cupones de bebé a su hija adolescente, lo que el hombre etiquetó como “una sugerencia a embarazarse” y la realidad es que su hija ocultó su embarazo hasta que Target lo “descubrió”.



Explicación

Desde la perspectiva de privacidad, los datos pueden clasificarse en cuatro categorías:

1. Información de identificación personal.	2. Cuasi-identificadores.	3. Información sensible.	4. Información no sensible.
Están asociados directamente con las personas: nombres, número de seguro social, correo electrónico, etcétera.	Datos que por sí solos pueden no ser útiles, pero al combinarse pueden ayudar a identificar al tipo de persona: género, código postal, edad, etcétera.	No son de identificación personal, pero sí es información delicada que en algún momento debe protegerse, como los ingresos económicos, información médica, de geolocalización, etcétera.	Cualquier otro tipo de información que no esté en cualquiera de las otras tres categorías.

Explicación

La **k-anonimidad** es una métrica de anonimato para la publicación de datos. Se obtiene a partir de un proceso llamado k-anonimización, mediante el cual se generalizan, modifican o distorsionan los valores de los cuasi-identificadores para que ningún individuo sea identificable de forma única de un grupo de k instancias dentro de un conjunto de datos (Slijepčević et al., 2021). El parámetro **k** indica el grado de anonimidad. El objetivo principal de este proceso es proteger la privacidad de los individuos cuyos datos se encuentran recopilados y almacenados.

Actividad

Universidad Tecmilenio

Busca por lo menos un ejemplo de la vida real en donde la privacidad de alguien o algunos se haya violado. Identifica las causas y efectos que provocó.

Cierre

Cada entidad que posee datos de particulares está sujeta a cumplir regulaciones sobre el tratamiento y uso que se les da a esos datos, y aunque algunas compañías afirman utilizar datos anónimos, las prácticas intrusivas que usan (envíos de ofertas, anuncios, etc.) sugieren que utilizan procesos de desanonimización para facilitar la reidentificación de las entidades, operando aparentemente al margen de las regulaciones.

Bibliografía

- Slijepčević, D., Henzl, M., Klausner, L., Dam, T., Kieseberg, P. y Zeppelzauer, M. (2021). k-Anonymity in practice: How generalisation and suppression affect machine learning classifiers. *Computers & Security, 111*(102488). Recuperado de https://www.sciencedirect.com/science/article/pii/S0167404821003126