

DATOS EN ADN

Samir Mateo Sierra Blanco

Universidad de Pamplona

Introducción a la ingeniería de sistemas

Fanny Casadiego Chiquillo

10 de diciembre de 2022

INTRODUCCIÓN

El presente trabajo busca definir y resaltar la importancia de la ciencia de datos, como también sus aplicaciones, tipos y sus pasos, además se habla sobre un ejemplo de aplicación de esto como lo es los datos en el ADN, se habla como se pueden guardar bits en secuencia de nucleótidos y realizar el proceso de conversión inverso, para realizar esto Microsoft ya ha desarrollado una máquina que puede realizar esa transformación, que revolucionará el almacenamiento de datos.

CIENCIA DE DATOS

La ciencia de datos es el estudio de datos con el fin de extraer información significativa para empresas. Es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos. Este análisis permite que los científicos de datos planteen y respondan a preguntas como “qué pasó”, “por qué pasó”, “qué pasará” y “qué se puede hacer con los resultados”.

¿Por qué es importante la ciencia de datos?

La ciencia de datos es importante porque combina herramientas, métodos y tecnología para generar significado a partir de los datos. Las organizaciones modernas están inundadas de datos; hay una proliferación de dispositivos que pueden recopilar y almacenar información de manera automática. Los sistemas en línea y los portales de pago capturan más datos en los campos del comercio electrónico, la medicina, las finanzas y cualquier otro aspecto de la vida humana. Disponemos de grandes cantidades de datos de texto, audio, video e imágenes.

Historia de la ciencia de datos

Aunque el término ciencia de datos no es nuevo, sus significados y connotaciones cambiaron con el tiempo. La palabra apareció por primera vez en los años 60 como nombre alternativo de la estadística. A finales de los 90, los profesionales de la computación formalizaron el término. Una propuesta de definición la consideraba un campo independiente con tres aspectos: diseño, recopilación y análisis de datos. Todavía tuvo que pasar otra década para que el término se utilizara fuera del ámbito académico.

Futuro de la ciencia de datos

La inteligencia artificial y las innovaciones del machine learning han hecho que el procesamiento de datos sea más rápido y eficiente. La demanda del sector ha creado un ecosistema de cursos, grados académicos y puestos de trabajo en el campo de la ciencia de datos. Debido al conjunto de competencias multidisciplinarias y a la experiencia necesaria, la ciencia de datos promete un fuerte crecimiento en las próximas décadas.

¿Para qué se utiliza la ciencia de datos?

La ciencia de datos se utiliza para estudiar los datos de cuatro maneras principales:

1. Análisis descriptivo

El análisis descriptivo examina los datos para obtener información sobre lo que ha ocurrido u ocurre en el entorno de datos. Se caracteriza por las visualizaciones de

datos, como los gráficos circulares, de barras o líneas, las tablas o las narraciones generadas.

2. Análisis de diagnóstico

El análisis de diagnóstico es un examen profundo o detallado de datos para entender por qué ha ocurrido algo. Se caracteriza por técnicas como el análisis detallado, el descubrimiento y la minería de datos o las correlaciones.

3. Análisis predictivo

El análisis predictivo utiliza los datos históricos para hacer previsiones precisas sobre los patrones de datos que pueden producirse en el futuro. Se caracteriza por técnicas como el machine learning, la previsión, la coincidencia de patrones y el modelado predictivo.

4. Análisis prescriptivo

El análisis prescriptivo lleva los datos predictivos al siguiente nivel. No solo predice lo que es probable que ocurra, sino que sugiere una respuesta óptima para ese resultado. Puede analizar las posibles implicaciones de las diferentes alternativas y recomendar el mejor curso de acción.

¿Cual es el proceso de la ciencia de datos?

Obtener datos

Los datos pueden ser preexistentes, recién adquiridos o un repositorio descargable de Internet. Los científicos de datos pueden extraerlos de las bases de datos internas o externas, del software CRM de la empresa, de los registros del servidor web, de las redes sociales o adquirirlos de terceros de confianza.

Depurar datos

La depuración o limpieza de datos consiste en el proceso de normalizarlos según un formato predeterminado. Incluye la gestión de los datos que faltan, la corrección de errores en estos y la eliminación de datos atípicos. Algunos ejemplos de la depuración de datos son:

- Cambiar todos los valores de fecha a un formato estándar común.
- Corregir las faltas de ortografía o los espacios adicionales.
- Corregir inexactitudes matemáticas o eliminar comas de números grandes.

Explorar datos

La exploración de datos es un análisis preliminar de estos que se utiliza para planificar otras estrategias para su modelado. Los científicos de datos obtienen una comprensión inicial de los datos mediante estadísticas descriptivas y herramientas de visualización de los mismos.

Modelar datos

El software y los algoritmos de machine learning se utilizan para obtener información más profunda, predecir resultados y prescribir el mejor curso de acción.

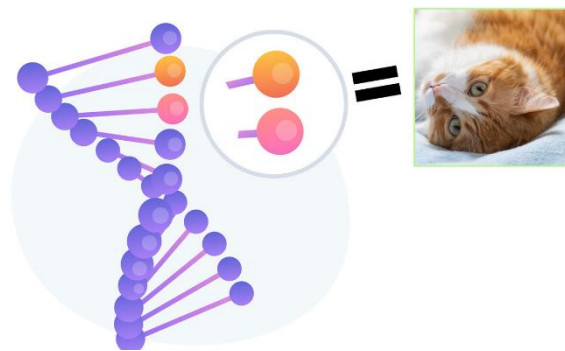
Interpretar los resultados

Los científicos de datos trabajan junto a los analistas y las empresas para convertir la información de datos en acción. Hacen diagramas, gráficos y tablas para representar tendencias y predicciones. La síntesis de datos ayuda a las partes interesadas a comprender y aplicar con eficacia los resultados.

DATOS EN ADN

Para almacenar algo digitalmente, como una foto, la convertimos en bits o códigos binarios. Cada pixel en una foto en blanco y negro es simplemente un cero o un uno. Y podemos escribir en el ADN de la misma manera que una impresora de inyección de tinta puede imprimir letras en una página: esos ceros y unos se convierten a A, T, C y G entonces lo escribimos, podemos almacenarlo y cuando queremos recuperar la fotografía, simplemente tendríamos que secuenciarlos.

El reto en realidad está en la tecnología y los dispositivos que harían estas operaciones. Las barreras financieras y de ingeniería para el almacenamiento viable de datos no genéticos en el ADN están aún en su infancia, superarlas provocaría una revolución en el almacenamiento y la seguridad, permitiendo que grandes cantidades de archivos se almacenen de forma segura en solo un gramo de materia.



Enlace de la exposición:

<https://drive.google.com/file/d/1PsvILFZ2hrIxsqPn5VdrXKVIGoCJwUE/view?usp=sharing>

BIBLIOGRAFIA

<https://aws.amazon.com/es/what-is/data-science/#:~:text=La%20ciencia%20de%20datos%20es,analizar%20grandes%20cantidades%20de%20datos>.

<https://www.computerweekly.com/es/definicion/Ciencia-de-datos>

<https://www.xataka.com/investigacion/nueva-tecnica-mit-para-almacenar-archivos-adn-mediante-capsulas-promete-hacer-facil-su-posterior-recuperacion>

<https://www.muycomputer.com/2022/03/07/almacenamiento-en-adn/>

<https://www.investigacionyciencia.es/noticias/logran-almacenar-datos-digitales-en-el-adn-de-bacterias-vivas-19437>

<https://www.nouschile.cl/almacenamiento-informacion-adn/>