Metodología Fundamental para la Ciencia de Datos

modelos para predecir resultados o para descubrir patrones subyacentes, con la meta de obtener insights **(Un insight es una clave, la esencia que nos permite encontrar la solución a un problema.)**

Analítica en las bases de datos relacionales y de Apache Hadoop. La analítica de texto en datos no estructurados o semiestructurados se está volviendo cada vez más importante como forma de incorporar a modelos predictivos la percepción y otra información útil de los textos, lo que a menudo conlleva mejoras significativas en la calidad y precisión del modelo.

**Esto nos deja ver que el análisis de datos nos esta llevando a mundo de intentar predecir el futuro a base de como se comportan los datos y los resultados de las estadísticas**

¿Qué es una metodología? Una metodología es una estrategia general que sirve de guía para los procesos y actividades que están dentro de un dominio determinado. La metodología no depende de tecnologías ni herramientas específicas, ni es un conjunto de técnicas o recetas. Más bien, la metodología proporciona al científico de datos un marco sobre cómo proceder con los métodos, procesos y argumentos que se utilizarán para obtener respuestas o resultados.

**La metodología es una herramienta guía para el científico que le ayuda a poder encontrar el camino correcto del uso de la información que le brindas una “x” de datos.**

**Diagrama

Descripción generada automáticamente**

Etapa 1: Comprensión del negocio:

Todos los proyectos comienzan con la comprensión del negocio.

**Esto se refiere comprender lo que el usuario desea como entrega final o como lo hemos entendido como REQUERIMIENTOS.**

Etapa 2: Enfoque analítico:

Cuando el problema empresarial se haya establecido claramente, el científico de datos podrá definir el enfoque analítico para resolver el problema.

**El analista brinda según sus conocimientos cual el cree que seria el mejor panorama que puede brindar a través de la idea que el usuario brinda en los requerimientos. Y si cumple con sus expectativas.**

Etapa 3: Requisitos de datos

El enfoque analítico elegido determina los requisitos de datos. Más concretamente, los métodos analíticos a utilizar requieren de determinados contenidos de datos, formatos y representaciones, orientados por el conocimiento en el dominio.

**Según conversamos creemos que es la parte donde se establece la forma en que se van a recopilar los datos o de donde van a provenir.**

Etapa 4: Recopilación de datos

En la etapa inicial de recopilación de datos, los científicos de datos identifican y reúnen los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del problema.

Al incorporar más datos, los modelos predictivos pueden representar mejor los eventos raros, como la incidencia de una enfermedad o un fallo del sistema.

**A la hora del análisis e intentar sacar estadísticas entre más historial pueden ser recopilados o cargados podríamos ver más escenarios reales.**

Etapa 5: Comprensión de datos

Después de la recopilación de datos inicial, los científicos de datos suelen utilizar estadísticas descriptivas y técnicas de visualización para comprender el contenido de los datos, evaluar su calidad y descubrir insights iniciales sobre ellos.

**Conversando creemos que los datos al inicio son solo eso datos los cuales debemos de trabajar y modelar para transfórmalos en información útil.**

Etapa 6: Preparación de datos

Esta etapa abarca todas las actividades para construir el conjunto de datos que se utilizará en la subsiguiente etapa de modelado.

La preparación de datos suele ser el paso más largo de los proyectos de ciencia de datos. En muchos dominios, algunos pasos de la preparación de datos son comunes para problemas diferentes. La automatización anticipada de determinados pasos de la preparación de datos puede acelerar el proceso al minimizar el tiempo de preparación a medida. Gracias al alto rendimiento, los sistemas masivamente paralelos y la funcionalidad analítica que reside donde se almacenan los datos de hoy en día, los científicos de datos pueden preparar los datos de forma más fácil y rápida utilizando conjuntos de datos muy grandes.

**El manejo de datos, validación que todos los formatos estén cargados de una forma correcta. Que la información sea analizada, comparada y sean cálculos exactos o lo más cercanos a eso.**

Etapa 7: Modelado:

La etapa de modelado utiliza la primera versión del conjunto de datos preparado y se enfoca en desarrollar modelos predictivos o descriptivos según el enfoque analítico previamente definido. En los modelos predictivos, los científicos de datos utilizan un conjunto de capacitación (datos históricos en los que se conoce el resultado de interés) para construir el modelo

**Esto es la parte que el científico hace ensayos para ver los escenarios que brindan los datos y ver cual información es la que realmente puede ser importante.**

Etapa 8: Evaluación Durante

El desarrollo del modelo y antes de su implementación, el científico de datos evalúa el modelo para comprender su calidad y garantizar que aborda el problema empresarial de manera adecuada y completa.

**Prueba y Error.**

Etapa 9: Implementación

Cuando el modelo satisfactorio ha sido desarrollado y aprobado por los promotores del negocio, se implementa en el entorno de producción o en un entorno de pruebas comparable.

**Esto es la prueba en campo en cierto sector de usuarios.**

Etapa 10: Retroalimentación

Al recopilar los resultados del modelo implementado, la organización obtiene retroalimentación sobre el rendimiento del modelo y su impacto en el entorno en el que se implementó. Por ejemplo, la retroalimentación puede ser en forma de porcentajes de respuesta a una campaña promocional dirigida a un grupo de clientes que ha sido identificado por el modelo como respondedores de alto potencial

**Ver el comportamiento del usuario y ver el porcentaje de satisfacción. O bien ver si se debe de actualizar algún requerimiento.**