



BIG DATA & DATA SCIENCE

PROGRAMA DE ESPECIALIZACIÓN



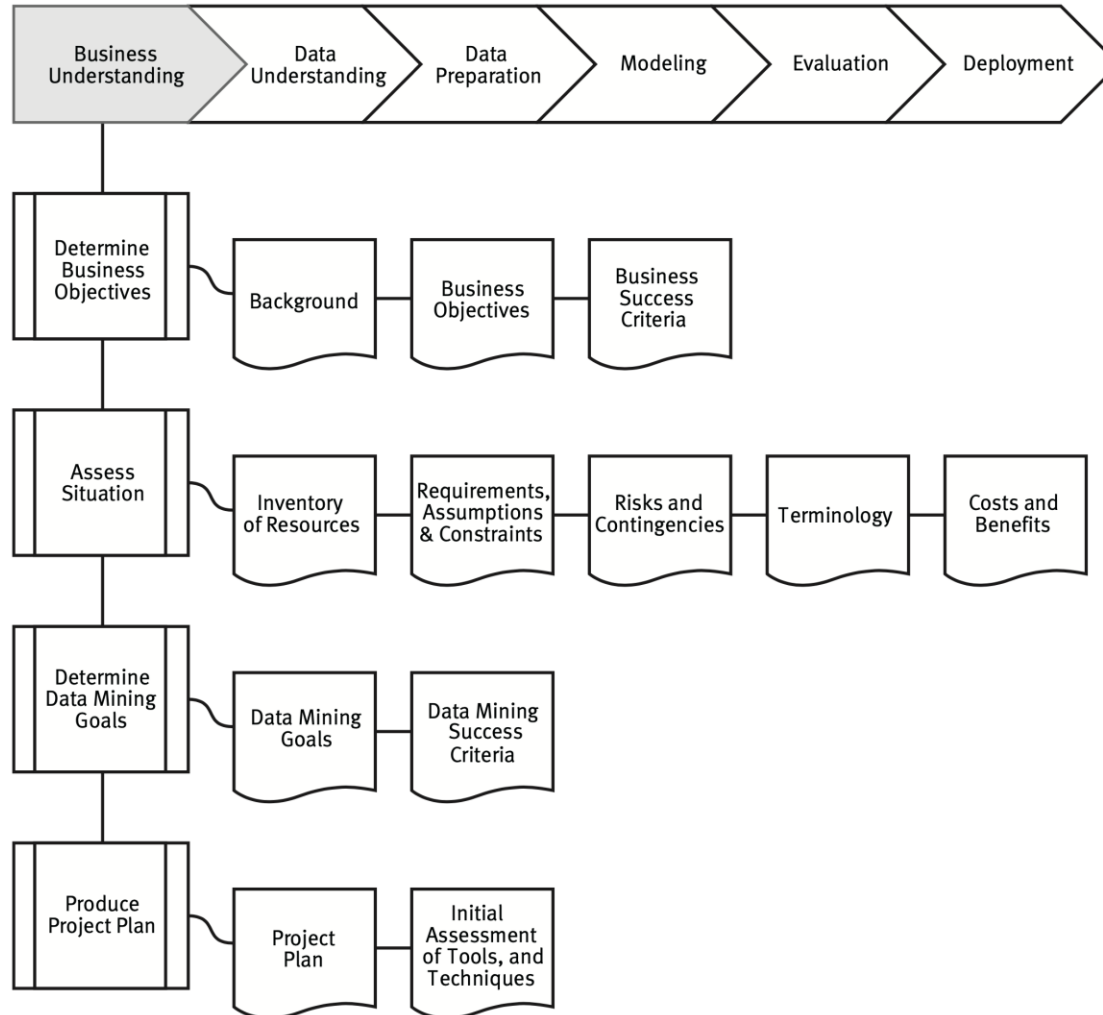
Business Data Discovery

Empezar con el negocio





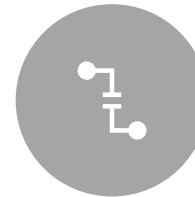
Entendimiento de Negocio CRISP-DM



Fase: Entendimiento de Negocio



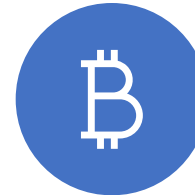
Esta fase inicial se enfoca en comprender los objetivos y requisitos del proyecto desde una perspectiva empresarial, en lugar de un punto de vista técnico.



El proceso de comprensión empresarial comienza con una idea de negocio, que se resuelve con una solución de Data & Analytics, que sienta las bases para la resolución exitosa del problema comercial.



Los patrocinadores de negocio que necesitan la solución analítica desempeñan un papel fundamental en esta etapa al definir el problema, los objetivos del proyecto y los requisitos de la solución desde una perspectiva comercial.



Cada proyecto comienzan con la comprensión del negocio, independientemente de su tamaño.



Para garantizar el éxito de cada proyecto, los promotores deben participar mientras dure el proyecto para proporcionar experiencia en el dominio empresarial, revisar los hallazgos intermedios y garantizar que el trabajo siga su curso para generar la solución deseada.



Se genera un plan de proyecto preliminar diseñado para lograr los objetivos

1. Determinación de los objetivos negocio



El primer objetivo es comprender a fondo, desde una perspectiva comercial, lo que el negocio realmente quiere lograr.



A menudo, el negocio tiene muchos objetivos competitivos y limitaciones que deben equilibrarse adecuadamente.



Se debe descubrir factores importantes al comienzo del proyecto que pueden influir en el resultado final.



Una consecuencia probable de descuidar este paso sería gastar un gran esfuerzo para producir las respuestas correctas a las preguntas incorrectas.

¿Cómo Determinar los Objetivos de Negocio?



Comenzar a recopilar información acerca de la situación comercial actual.



Registrar los objetivos comerciales específicos que decidan los gerentes.



Consensuar los criterios que se utilizarán para determinar el rendimiento del proceso desde una perspectiva comercial.



Antecedentes



Recopilar la información que se conoce sobre la situación comercial de la organización al comienzo del proyecto.



Sirven para identificar y conocer más de cerca los objetivos comerciales que se deben alcanzar.

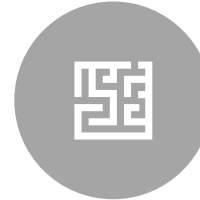


Sirven para identificar los recursos, tanto humanos como materiales, que se pueden usar o necesitar durante el curso del proyecto.

Objetivos de negocio



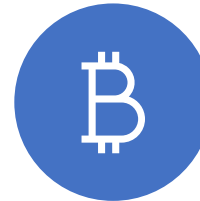
Describir el objetivo principal del negocio, desde una perspectiva negocio.



Describe informalmente el problema a resolver.



Especificar todas las preguntas comerciales con la mayor precisión posible.



Especificar cualquier otro requisito de negocio (por ejemplo, la empresa no quiere perder ningún cliente).



Especificar los beneficios esperados en términos de negocio (comerciales, operativos, etc.).



Tener cuidado de establecer objetivos inalcanzables: hacerlos lo más realistas posible.

Objetivos de negocio



Asignación de
recursos



Segmentación
de clientes



Benchmarking
comparativo



Orientado al
cliente

Asignación de recursos

Previsiones precisas para asignar recursos
tales como:



Personal



Presupuesto



Inventario para obtener un
resultado de calidad y
mantener los costos bajos.

Segmentación de clientes



Identificar segmentos de clientes más granulares para personalizar la experiencia de estos y, por lo tanto, mejorar las relaciones de conversión.



Clientes potenciales.



Descontentos o felices.

Benchmarking comparativo



Crear puntos de referencia comparativos en un grupo de individuos o grupos de trabajo para encontrar valores atípicos positivos y negativos.



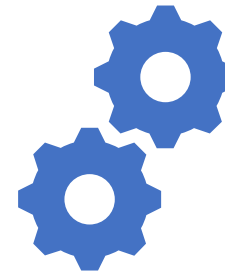
Aprender de lo positivo y remediar lo negativo.

Orientado al cliente

Todas las relaciones se basan en la confianza,
especialmente las relaciones comerciales.



Externalizar la analítica a los clientes,
socios y proveedores.



Incrustar análisis en procesos
comerciales externos para crear
rigidez.

Criterios de éxito empresarial



Describir los criterios para un resultado exitoso o útil para el proyecto desde el punto de vista empresarial.



Documentar los criterios de rendimiento del proyecto.



Asegurar de que cada objetivo comercial tiene un criterio de rendimiento relacionado.



Especificar los criterios de éxito empresarial (por ejemplo, mejorar la tasa de respuesta en una campaña de correo en un 10 por ciento).



Identificar quién evalúa los criterios de éxito.



Necesidad	Caso de uso	Beneficios
Customer Relation Procesar solicitudes de clientes de primer nivel. Ayudar a los empleados a diario con tareas repetitivas o que consumen mucho tiempo.	<div data-bbox="645 368 807 411"></div> <div data-bbox="821 361 1251 432">Air France ha lanzado su chatbot "Louis" para recuperar equipaje.</div> <div data-bbox="678 468 770 561"></div> <div data-bbox="821 475 1309 546">AXA utiliza el aprendizaje automático para analizar el motivo de los envíos</div>	<div data-bbox="1367 361 1763 432">Oferta de servicio al cliente automatizado disponible 24/7</div> <div data-bbox="1367 475 1854 582">Aumentar la productividad mediante la automatización de tareas administrativas.</div>
Marketing Automatización de procesamiento masivo de datos.	<div data-bbox="683 675 801 718"></div> <div data-bbox="821 646 1290 753">Netflix usa IA en su plataforma para ofrecer contenido específico a sus suscriptores.</div> <div data-bbox="683 796 801 896"></div> <div data-bbox="821 796 1319 903">Toyota utiliza paneles artificiales que automatizan y agregan la recopilación de datos del consumidor.</div>	<div data-bbox="1367 646 1850 718">Mejorar la experiencia del cliente (personalizaciones, recomendación).</div> <div data-bbox="1367 761 1870 832">Desarrollar un mejor conocimiento de la base de clientes.</div> <div data-bbox="1367 875 1831 946">Mejorar la satisfacción del cliente y LVT.</div>
Retail Implementar cámaras y algoritmos avanzados que rastreen automáticamente a compradores y productos.	<div data-bbox="649 1032 811 1075"></div> <div data-bbox="821 1003 1329 1189">Walmart ha desarrollado una tecnología que permite compras Grab-and-go para facilitar el proceso de compra del cliente en el punto de venta físico.</div>	<div data-bbox="1367 1003 1831 1075">Facilitar el pago que ya no requiere una transacción física.</div> <div data-bbox="1367 1118 1850 1268">Identificar y explotar los datos de los clientes (preferencias, comportamiento) para abordarlos mejor.</div>

Caso de uso - Antecedentes

- Una empresa de ventas en la Web está experimentando un crecimiento en la competencia de nuevos sitios.
- Al afrontar la realidad de que las tiendas Web están compitiendo por atraer más clientes y tener más leads, la empresa debe encontrar fórmulas para mantener la rentabilidad, sin aumentar el coste de la adquisición de clientes.
- Una solución propuesta es cultivar las relaciones de los clientes actuales con objeto de aumentar al máximo el valor de cada uno de estos.

Caso de uso - Objetivos de negocio



Mejorar las ventas cruzadas realizando mejores recomendaciones.



Fomentar la lealtad de los clientes con un servicio más personalizado.



Caso de USO- Criterios de éxito empresarial



Las ventas cruzadas se aumentan en un 10%.



Los clientes pasan más tiempo y ven más páginas en el sitio por cada visita.



El proyecto se complete dentro del plazo y del presupuesto.

2. Valoración de la situación

Implica una búsqueda de hechos más detallada sobre todos los recursos, limitaciones, suposiciones y otros factores que deben considerarse al determinar el objetivo del análisis de datos y al desarrollar el plan del proyecto.



Realizar una
valoración de su
situación actual



¿Dispone del
personal
necesario para
completar el
proyecto?



¿Qué tipos de
datos están
disponibles
para el
análisis?



¿Cuáles son los
principales
factores de
riesgo?



¿Dispone de
planes de
contingencia
para cada factor
de riesgo?

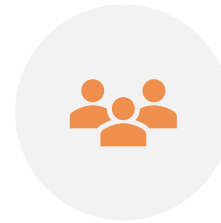
Inventario de recursos



VALORAR LOS
RECURSOS DE
HARDWARE



IDENTIFICAR
ORÍGENES DE
DATOS Y
ALMACENES DE
CONOCIMIENTOS



IDENTIFICAR
PERSONAL



ES INDISPENSABLE
MANTENER UN
INVENTARIO
ACTUALIZADO DE
LOS RECURSOS

Requisitos, suposiciones y restricciones



Determinar requisitos

¿Existen restricciones legales y de seguridad sobre los datos?

¿Los usuarios están alineados con los requisitos de planificación del proyecto?



Describir los supuestos

¿Existen factores económicos que pueden afectar al proyecto?

¿Existen supuestos de calidad de datos?



Comprobar las restricciones

¿Dispone de todas las contraseñas necesarias para acceder a los datos?

¿Las restricciones financieras están incluidas en el presupuesto del proyecto?

Riesgos y contingencias



Identificar riesgos

- Comerciales, organizacionales, datos, financieros, técnicos, etc.



Desarrollar planes de contingencia

- Determinar las condiciones bajo las cuales puede ocurrir cada riesgo
- Desarrollar planes de contingencia

Terminología



Compile un glosario de terminología relevante para el proyecto.



Un glosario de terminología comercial relevante, que forma parte de la comprensión comercial disponible para el proyecto.



Un glosario de terminología analítica, ilustrado con ejemplos relevantes para el problema comercial en cuestión.

Costos y beneficios



Costos
estimados para
la recolección
de datos



Estimar los costos de
desarrollar e
implementar una
solución



Identificar los
beneficios (mejor
satisfacción del cliente,
ROI y aumento de los
ingresos)



Estimar los costos
operativos.



¿Qué es la inteligencia Artificial?

La Inteligencia Artificial (IA) se puede definir como el uso de algoritmos informáticos para intentar la simulación de la inteligencia humana en algunas de sus dimensiones, como el procesamiento de las entradas de nuestros sentidos: visión (visión por computadora), audio (reconocimiento de voz), lenguaje (procesamiento del lenguaje natural), o habilidades de reconocimiento de patrones (aprendizaje automático), entre otros.



Machine Learning
(Learn)



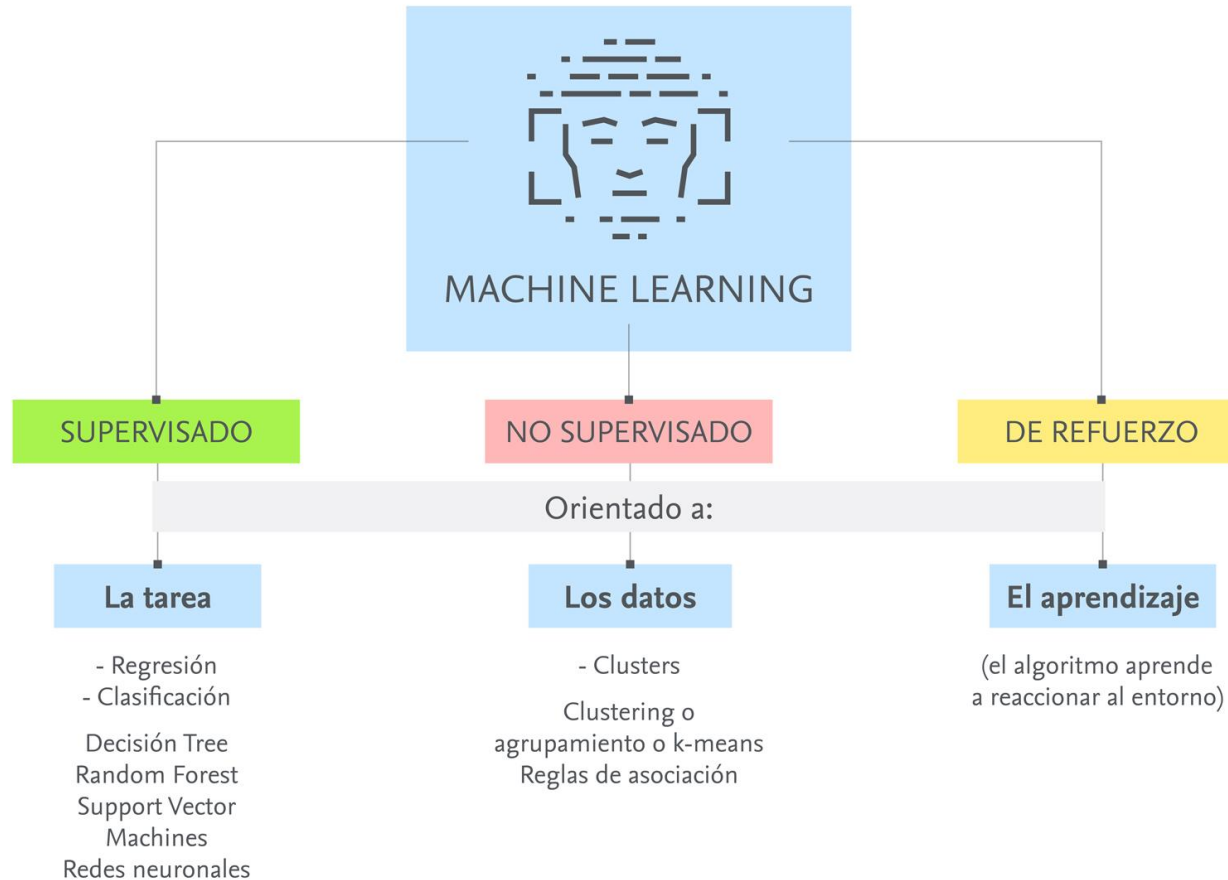
Computer Vision
(See)



Natural Language Processing (NLP)
(Communicate)



Speech Recognition
(Hear)





Aprendizaje supervisado

Donde el algoritmo se enfoca en hacer una predicción precisa de un indicador específico, que dependiendo del tipo de indicador puede convertirse en:

- **Regresión:** predicción de un indicador continuo, es decir, un indicador que puede tener muchos valores posibles, incluso a nivel decimal. Un ejemplo es la previsión de la demanda, que puede tener una amplia gama de posibles valores previstos.
- **Clasificación:** predicción de un indicador discreto, es decir, un indicador que tiene un par de valores posibles. Por ejemplo, la predicción de la rotación de clientes solo necesita considerar dos escenarios posibles (abandono vs no cancelado).

Evaluación del modelo: la bondad de ajuste de los algoritmos supervisados se puede medir por sus métricas de precisión y se puede mejorar continuamente mediante la infusión de más datos.



Aprendizaje no supervisado

El algoritmo se enfoca en hacer una clasificación óptima de un conjunto de datos completo, que dependiendo del tipo de clasificación que se necesite puede convertirse en:

- **Clustering:** clasificación de filas (unidades, personas) en el conjunto de datos. Por ejemplo, un algoritmo de segmentación de clientes clasificaría a todas las personas incluidas en un conjunto de datos determinado.
- **Reducción de dimensionalidad:** clasificación de columnas (variables, características) en el conjunto de datos. Por ejemplo, al analizar una gran cantidad de indicadores digitales, el investigador podría estar interesado en buscar una versión reducida del mismo y generar un “índice de digitalización”.

Evaluación del modelo: los algoritmos no supervisados carecen de métricas de precisión, por lo que necesitan interpretación humana para garantizar que la clasificación propuesta esté alineada con las necesidades comerciales.



3.

Determinación de los objetivos de minería de datos

- Una meta comercial establece objetivos en terminología comercial; una meta de minería de datos establece los objetivos del proyecto en términos técnicos.
- Por ejemplo, el objetivo comercial podría ser: "Aumentar las ventas de catálogos a clientes existentes", mientras que un objetivo de minería de datos podría ser: "Predecir cuántos widgets comprará un cliente, dadas sus compras en los últimos tres años, información demográfica relevante y el precio del artículo ".

Objetivos de minería de datos



Cuando se trabaja con analistas de negocio y de datos para definir una solución técnica al problema comercial, es necesario recordar que los objetivos deben ser concretos.



Traducir preguntas comerciales a objetivos de minería de datos.



Especifique el tipo de problema de minería de datos (clasificación, descriptivo, predicción y clustering).



De ser posible, proporcione datos reales para resultados deseados.

Criterios de éxito de minería de datos



Especificar criterios para la evaluación del modelo (por ejemplo, precisión del modelo, rendimiento y complejidad).



Definir puntos de referencia para los criterios de evaluación.



Especifique los criterios que abordan los criterios de evaluación subjetiva (por ejemplo, la capacidad de explicación del modelo y los datos, y la información de marketing proporcionada por el modelo).



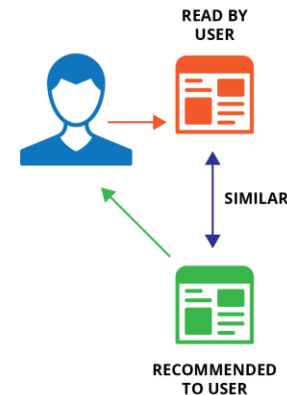
Los criterios de éxito de Analytics son diferentes a los criterios de éxito empresarial definidos anteriormente.



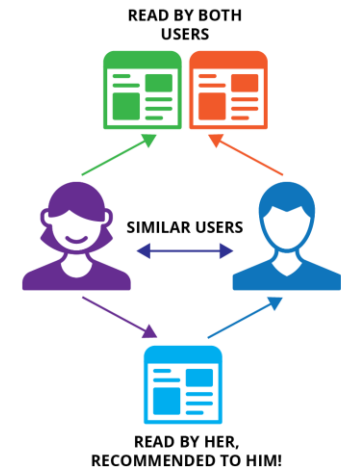
Motores de Recomendación

- Los **Motores de Recomendación** también conocidos como **Sistemas de Recomendación**, ayudando a los usuarios para filtrar toneladas de elementos que ellos no desean. Se utilizan en prácticamente todos los servicios populares para reducir la carga de elección desde anuncios de Instagram hasta sugerencias de productos de Amazon. Los motores de recomendación aumentan la posibilidad de una conversión al devolver los resultados correctos a un usuario. Existen 2 tipos de motores de recomendación
 - Filtrado colaborativo
 - Filtrado de contenido

CONTENT-BASED FILTERING



COLLABORATIVE FILTERING

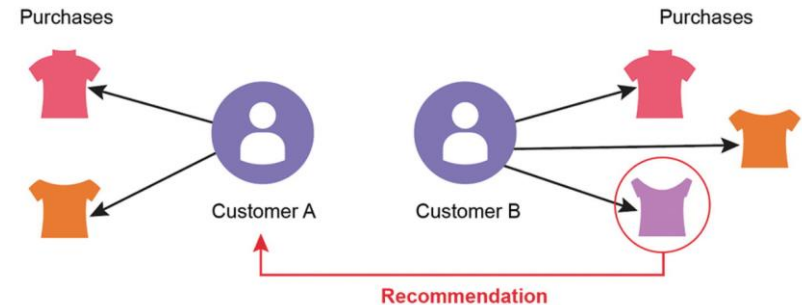




Filtrado Colaborativo

El **Filtrado Colaborativo** utiliza información de un gran conjunto de datos de compras de usuarios y otros comportamientos para predecir lo que busca otro cliente. Hay dos enfoques para el filtrado colaborativo: un enfoque basado en el usuario y un enfoque basado en elementos.

- **Basado en Usuario:** Devuelve recomendaciones basadas en la similitud de un usuario con otros usuarios. Ejemplo, incluye recomendaciones de productos basadas en lo que otros usuarios han comprado.
- **Basado en Elementos:** Recomienda productos en función de cómo los usuarios calificaron elementos similares. Éste es útil porque puede brindar recomendaciones relevantes incluso si no sabemos nada sobre un usuario determinado.

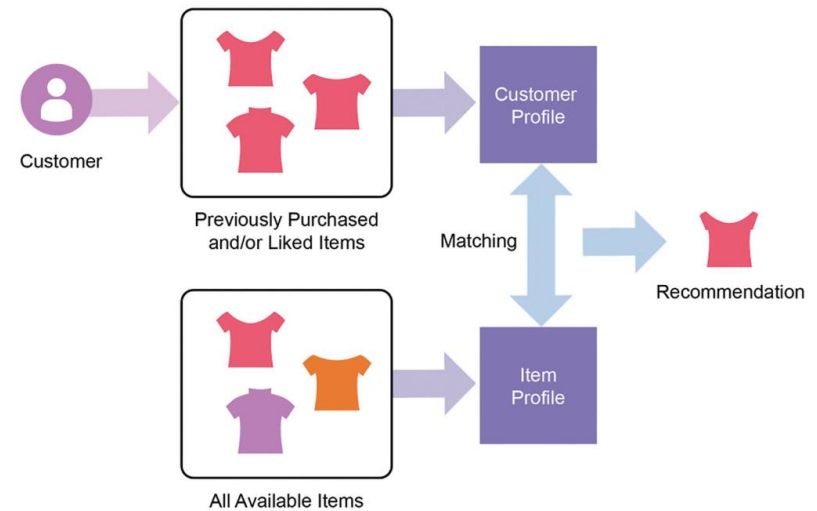




Filtrado Basado en Contenido

Los métodos de Filtrado Basado en Contenido se basan en las acciones y preferencias del usuario.

Si un usuario está explorando un sitio y le gusta comprar sólo vestidos rojos entonces se le revelarán más vestidos rojos durante su búsqueda punto sin embargo como un problema con este método es que pueden continuar recomendando productos en la misma categoría potencialmente incluso después de que el usuario haya perdido interés en buscar artículos en esta categoría.





4. Producción de un plan de proyecto



Describir el plan previsto para lograr los objetivos del proyecto de ciencia de datos y alcanzar los objetivos comerciales.



Informar a todos los usuarios relacionados con los objetivos, recursos, riesgos del proyecto y programar todas las fases.

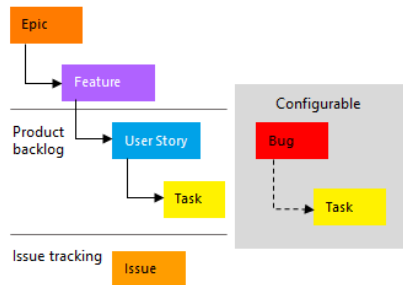


Listar las etapas que se ejecutarán en el proyecto, junto con su duración, recursos necesarios, entradas, salidas y dependencias.

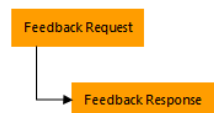


Plan de proyecto Vs Procesos Ágil

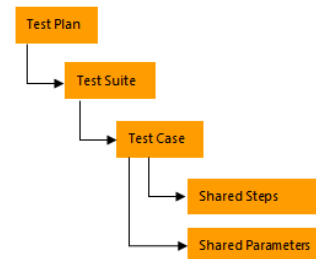
Portfolio backlog



Feedback



Test



Code review

