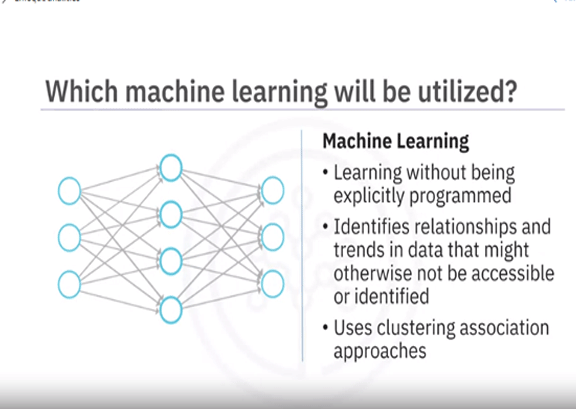
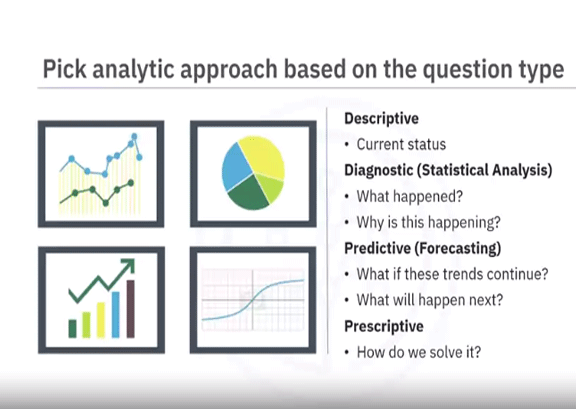
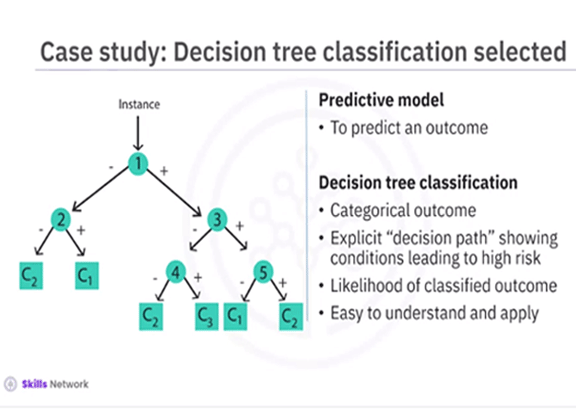
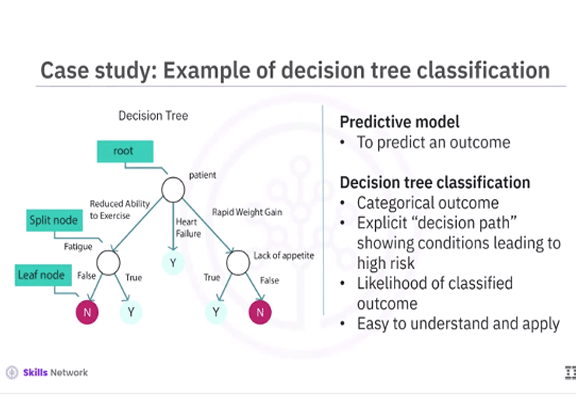


Enfoque analitico









Desde los requerimientos a la recopilacion de datos

Basándose en la comprensión del problema en cuestión, y luego utilizando el enfoque analítico seleccionado, el Data Scientist está listo para comenzar.

En el estudio de caso, la primera tarea fue definir los requisitos de datos para el

enfoque de clasificación del árbol de decisiones seleccionado.

Esto incluyó la selección de una cohorte de pacientes adecuada de la

base de afiliados de proveedores de seguros médicos. Con el fin de compilar las historias clínicas completas, se identificaron tres criterios para su inclusión en la cohorte.

Primero, un paciente necesitaba ser ingresado como paciente internado dentro del área de servicio del proveedor, para que tuviera acceso a la información necesaria.

En segundo lugar, se centraron en pacientes con diagnóstico primario de insuficiencia cardíaca congestiva durante un año completo.

En tercer lugar, el paciente debe haber tenido una matrícula continua durante al menos seis meses, antes del ingreso primario por insuficiencia cardíaca congestiva, para

poder compilar la historia clínica completa.

Los pacientes con insuficiencia cardíaca congestiva que también habían sido diagnosticados con otras afecciones médicas significativas, fueron excluidos de la cohorte porque estas afecciones provocarían tasas de reingreso superiores a la media y, por lo tanto, podrían sesgar los resultados.

Luego se definió el contenido, el formato y las representaciones de los datos necesarios para la clasificación del árbol de decisiones.

Esta técnica de modelado requiere un registro por paciente, con columnas que representan

las variables del modelo.

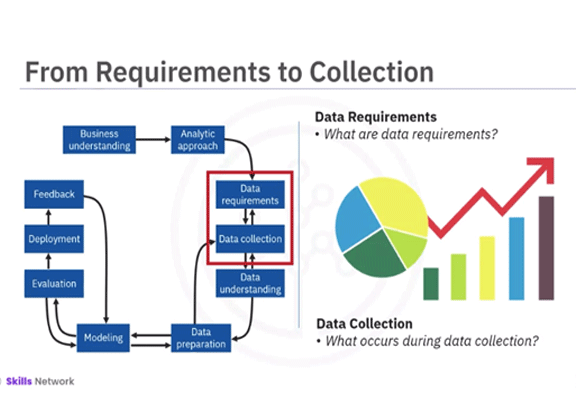
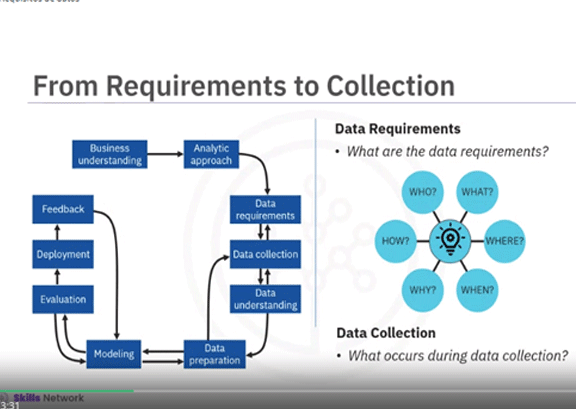
Para modelar el resultado de la readmisión, era necesario contar con datos que cubran todos los aspectos de la historia clínica del paciente.

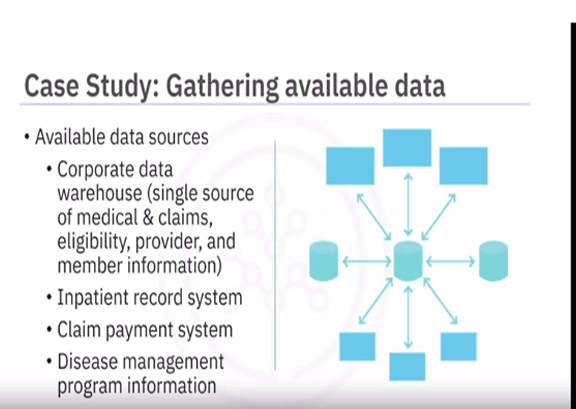
Este contenido incluiría admisiones, diagnósticos primarios, secundarios y terciarios, procedimientos, prescripciones y otros servicios prestados durante la hospitalización o durante las visitas al paciente o al médico.

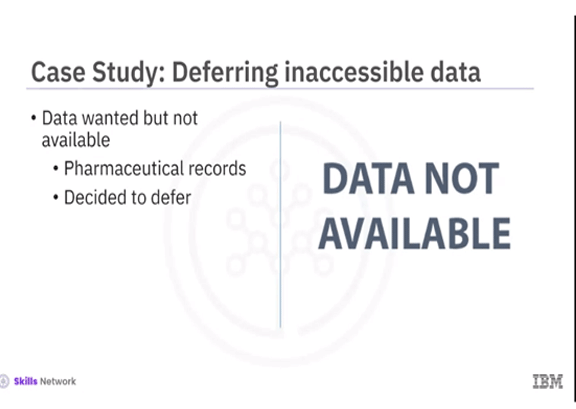
Por lo tanto, un paciente en particular podría tener miles de registros, representando todos sus atributos relacionados.

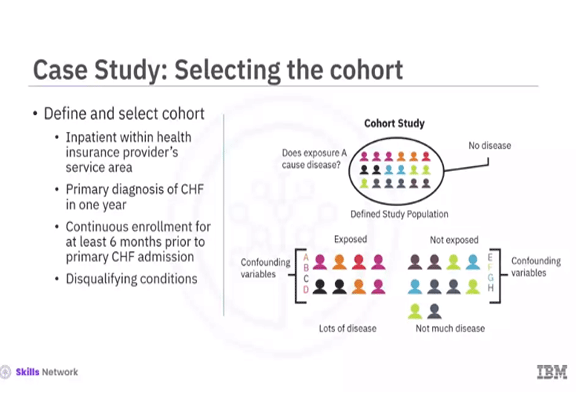
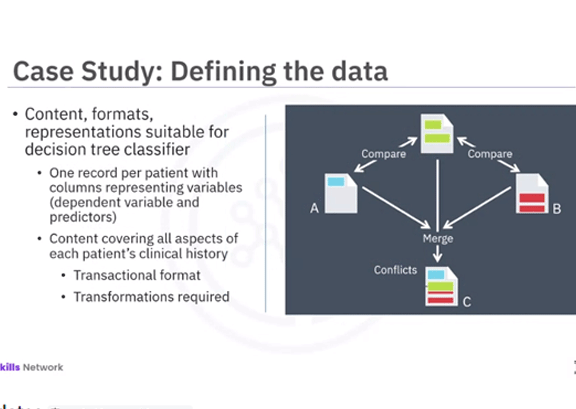
Para llegar al formato de un registro por paciente, los científicos de datos enrollaron los

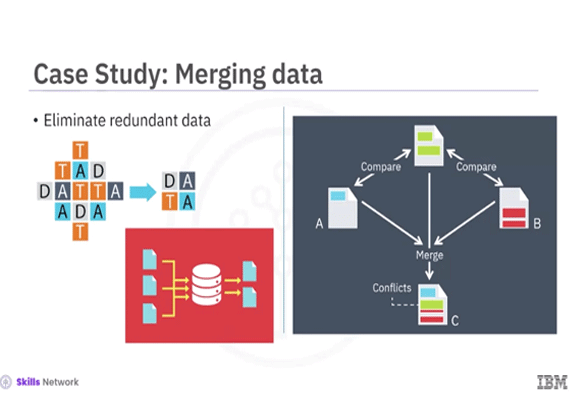
registros transaccionales al nivel del paciente, creando una serie de nuevas variables para representar esa información.











En esta fase se revisan los requerimientos de datos y se decide si la recolección

requiere más, o menos datos.

Una vez recolectados los ingredientes de datos en la etapa de recolección, el científico de datos tendrá una buena comprensión de los insumos con los cuales trabajará.

Técnicas como la estadística descriptiva y la visualización pueden aplicarse al conjunto

de datos para evaluar su contenido, calidad y conclusiones iniciales.

Se identificarán vacíos en los datos y se harán planes, ya sea para llenarlos o

hacer sustituciones.

En esencia, los ingredientes están ahora sobre la tabla de corte.

Ahora veamos algunos ejemplos de la etapa de recolección de la metodología de la ciencia de datos. Esta etapa se emprende al finalizar la etapa de requerimientos de datos.

Veamos el estudio de caso en relación con la aplicación de la "Recolección de datos".

Para recolectar datos, necesitas saber la fuente o el modo de hallar los elementos de datos

que se requieren. En el contexto de nuestro estudio de caso, estos pueden incluir:

información demográfica, clínica o de cobertura de los pacientes,

información de proveedor, registros de reclamos, al igual que

información farmacéutica y adicional relacionada con diagnósticos de pacientes de

insuficiencia cardíaca. En este estudio de caso, se necesitaba cierta información de medicamentos, pero esa fuente de datos aún no estaba integrada al resto de las fuentes de datos. Esto nos lleva a un punto importante: Está bien aplazar decisiones acerca de datos no disponibles, e intentar adquirirlos en una etapa posterior.

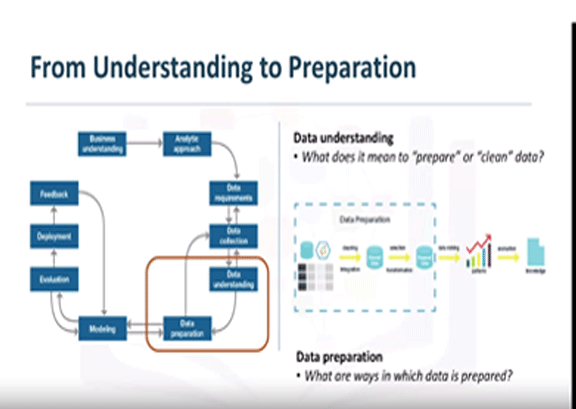
Por ejemplo, esto puede hacerse aún después de obtener resultados intermedios del modelado

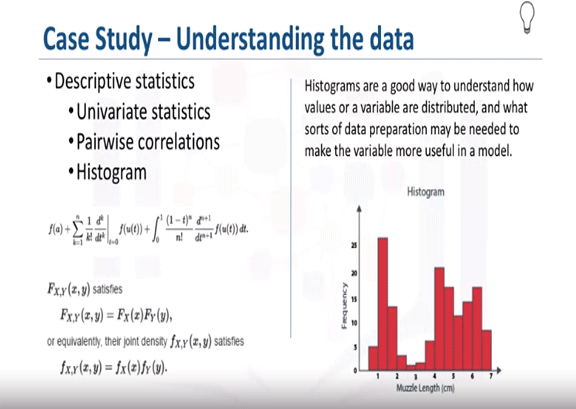
predictivo.

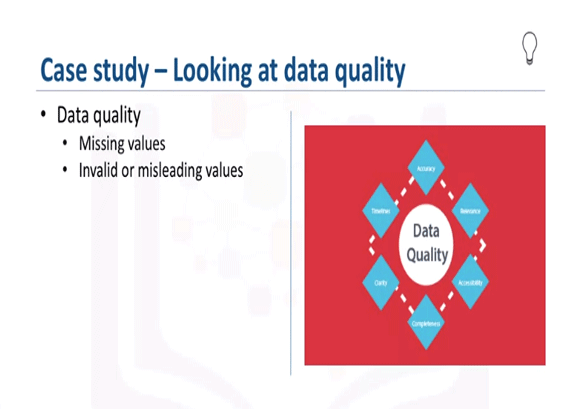
Si esos resultados indican que la información de medicamentos será importante para lograr un buen modelo,

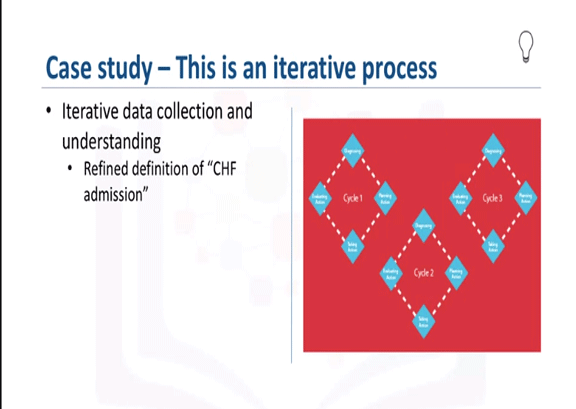
se invertiría el tiempo para intentar conseguirla.

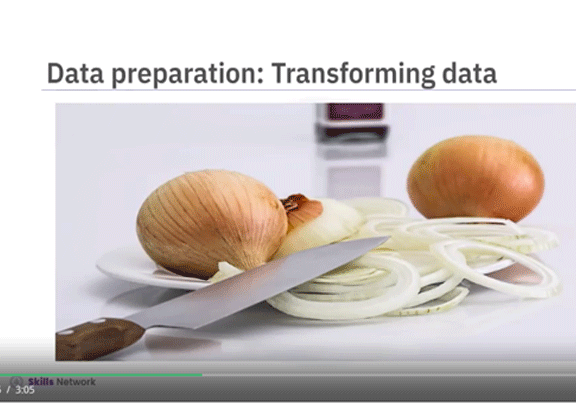
Comprender los datos











En el estudio de caso, la primera tarea fue definir los requisitos de datos para el

enfoque de clasificación del árbol de decisiones seleccionado.

Esto incluyó la selección de una cohorte de pacientes adecuada de la

base de afiliados de proveedores de seguros médicos. Con el

fin de compilar las historias clínicas completas, se identificaron tres criterios

para su inclusión en la cohorte.

Primero, un paciente necesitaba ser ingresado como paciente internado dentro del área de servicio del proveedor, para

que tuviera acceso a la información necesaria.

En segundo lugar, se centraron en pacientes con diagnóstico primario de insuficiencia cardíaca congestiva durante

un año completo.

En tercer lugar, el paciente debe haber tenido una matrícula continua durante al menos seis meses, antes

del ingreso primario por insuficiencia cardíaca congestiva, para

poder compilar la historia clínica completa.

Los pacientes con insuficiencia cardíaca congestiva que también habían sido diagnosticados con otras

afecciones médicas significativas, fueron excluidos de la cohorte porque estas afecciones provocarían

tasas de reingreso superiores a la media y, por lo tanto, podrían sesgar los resultados.

Luego se

definió el contenido, el formato y las representaciones de los datos necesarios para la clasificación del árbol de decisiones.

Esta técnica de modelado requiere un registro por paciente, con columnas que representan

las variables del modelo.

Para modelar el resultado de la readmisión, era necesario contar con datos que cubran todos los aspectos de la

historia clínica del paciente.

Este contenido incluiría admisiones, diagnósticos primarios, secundarios y terciarios, procedimientos,

prescripciones y otros servicios prestados durante la hospitalización o durante las

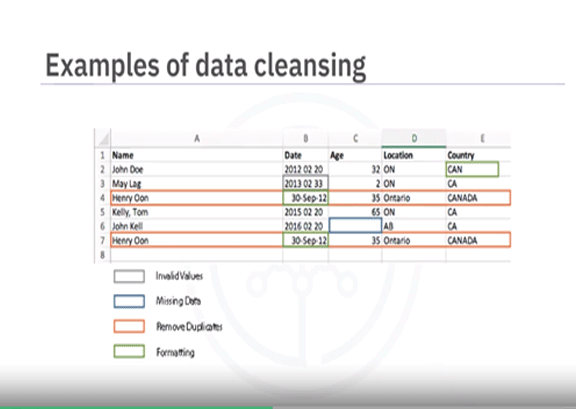
visitas al paciente o al médico.

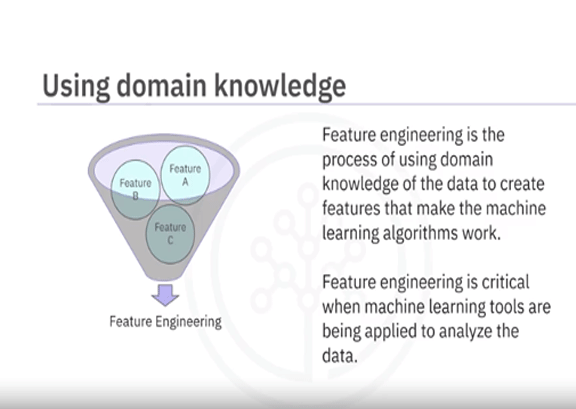
Por lo tanto, un paciente en particular podría tener miles de registros, representando todos sus

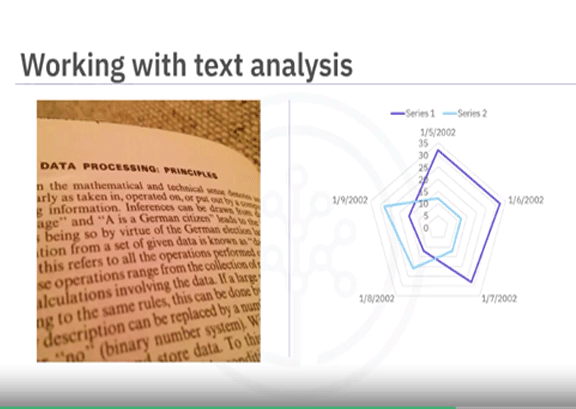
atributos relacionados.

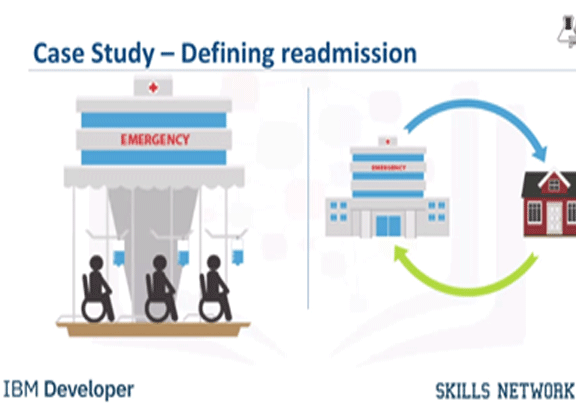
Para llegar al formato de un registro por paciente, los científicos de datos enrollaron los

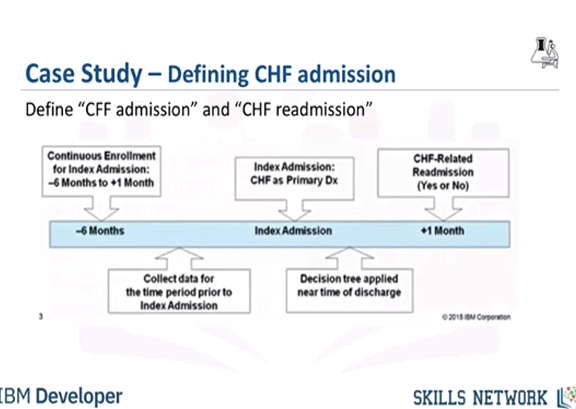
registros transaccionales al nivel del paciente, creando una

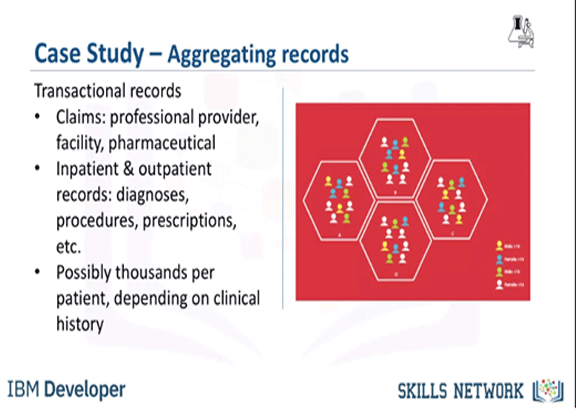


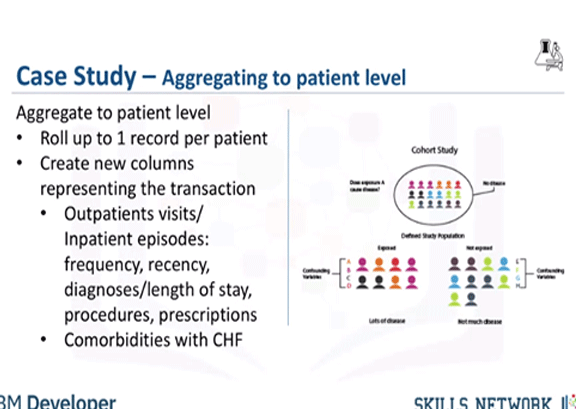


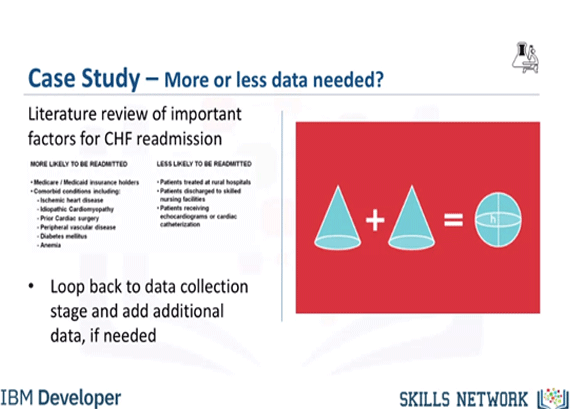


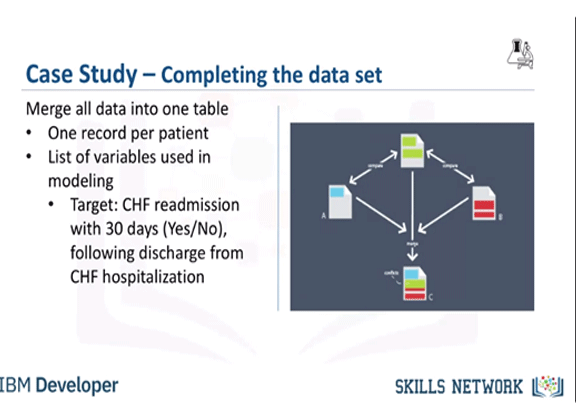


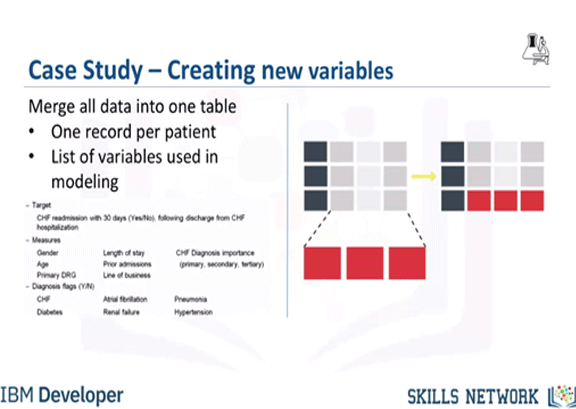


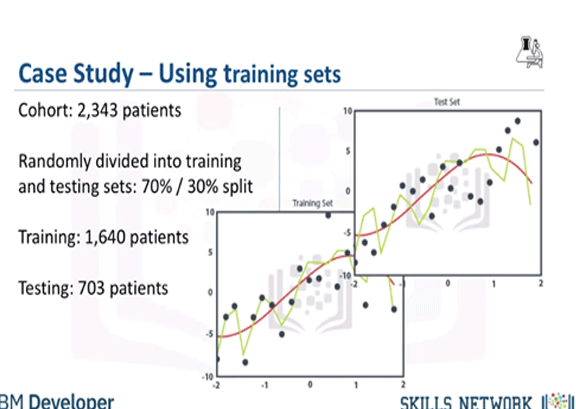






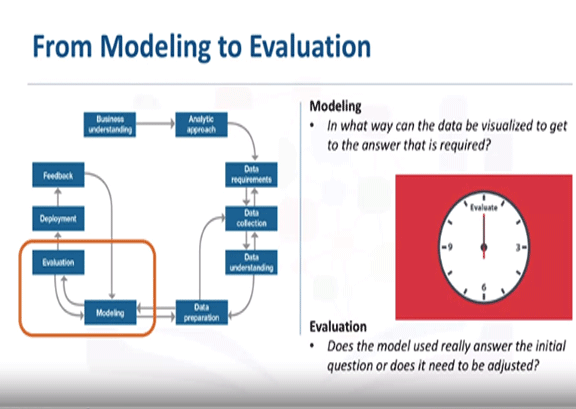


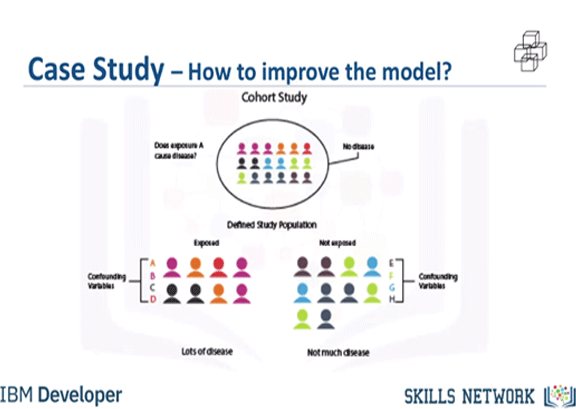


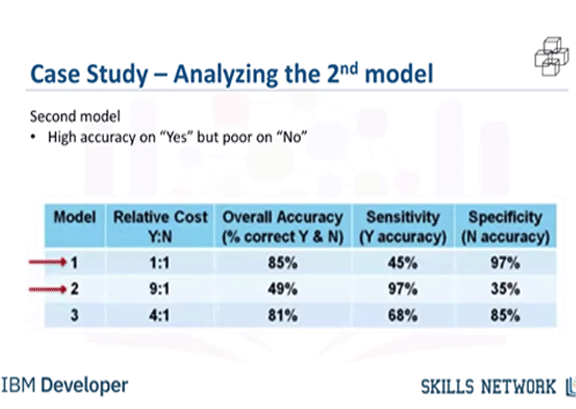


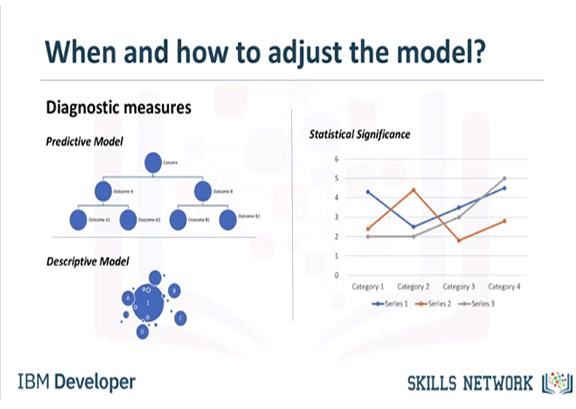
**Modelado**

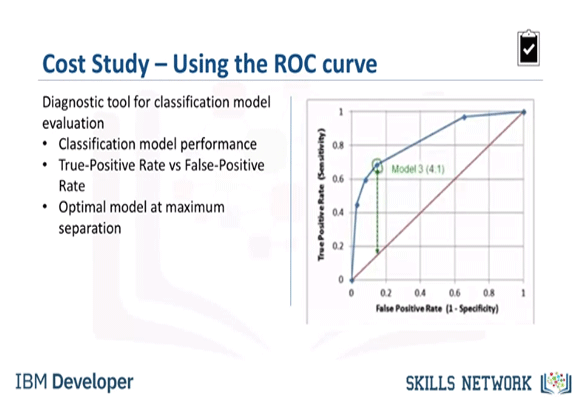
Aquí discutiremos uno de los muchos aspectos de la construcción de modelos, en este caso, el ajuste de parámetros para mejorar el modelo. Con un conjunto de entrenamiento preparado, se puede crear el primer modelo de clasificación de árbol de decisión para el reingreso por insuficiencia cardíaca congestiva. Buscamos pacientes con alto riesgo de reingreso, por lo que el resultado de interés será el reingreso por insuficiencia cardíaca congestiva "Sí/=Sí". En este modelo inicial, la precisión general al clasificar los resultados de sí y no fue del 85%. Esto suena bien, pero sólo representa el 45% de los “sí”. Los reingresos reales se clasificaron correctamente, lo que significa que el modelo no es muy preciso. La pregunta entonces es: ¿Cómo se puede mejorar la precisión del modelo para predecir el resultado Sí? El mejor parámetro a establecer para la clasificación del árbol de decisión es el de clasificación errónea. es el costo relativo de los resultados de sí y no. Piénselo de esta manera: cuando una verdadera falta de reingreso se clasifica erróneamente y se toman medidas para reducir el riesgo de ese paciente, el costo de ese error es una intervención desperdiciada. Un estadístico llama a esto error de tipo I o falso positivo. Pero cuando un verdadero reingreso se clasifica erróneamente y no se toman medidas para reducir este riesgo, entonces el costo de este error es el reingreso y todos los costos asociados más el trauma para el paciente. Este es un error de tipo II o falso negativo. Por lo tanto, los costos de dos errores de clasificación diferentes tendrán algunas características. Por lo tanto, tiene sentido ajustar los pesos relativos de clasificar erróneamente los resultados de sí y no. El valor predeterminado es 1 a 1, pero el algoritmo del árbol de decisión permite establecer un valor más alto para sí. Se determinó que el costo relativo para el segundo modelo era de 9 a 1. Esta es una proporción muy alta, pero da más información sobre el comportamiento del modelo. Esta vez el modelo clasificó correctamente el 97% de los sí, pero a expensas de una precisión muy baja en los no, con una precisión general de sólo el 49%. Claramente este no era un buen modelo. El problema con este resultado es la gran cantidad de falsos positivos que sugerirían una intervención innecesaria y costosa para los pacientes que no serían readmitidos. Por lo tanto, el científico de datos debe intentar nuevamente encontrar un mejor equilibrio entre precisión de sí y no. Para el tercer modelo, el coste relativo se fijó en un 4 a 1 más razonable. Esta vez, se logró un 68% de precisión solo en el caso de sí, lo que los estadísticos llaman sensibilidad, y un 85% de precisión en el no, llamado especificidad, con un 81% de precisión en general. Esta es la mejor compensación que se puede lograr con un conjunto de entrenamiento bastante pequeño ajustando el costo relativo de clasificar erróneamente el parámetro de resultado de sí y no. Por supuesto, se requiere mucho más trabajo para el modelado, incluida la iteración de regreso a la fase de preparación de datos para redefinir algunas de las otras variables, representar mejor la información subyacente y así mejorar el modelo

****

****

****

****

****

**Despliegue**

Al prepararse para la implementación de la solución, el siguiente paso es reducir el riesgo de readopción.para el grupo de trabajo que diseñará y gestionará el programa de intervención.

era absorber la información.

En este escenario, los empresarios se preguntan cómo identificará el personal de la clínica a los pacientes de alto riesgo.

y modelar resultados para comprender cómo diseñar acciones de respuesta apropiadas.

traducido.

Por supuesto, el objetivo es que estos pacientes reingresen dentro de los 30 días posteriores al alta.

era reducir la probabilidad de aceptación.

Durante la fase de requisitos comerciales, el Director del Programa de Intervención y su equipo,

Riesgo automático, casi en tiempo real, de insuficiencia cardíaca congestiva.

Quería una aplicación que les permitiera evaluar.

Una aplicación basada en tableta que todos los miembros del personal pueden llevar consigo.

Sus datos fueron creados durante su estancia en el hospital.

Estos datos de los pacientes se generaron durante la estancia hospitalaria.

Se preparará automáticamente en el formato que necesite el modelo y cada paciente.

se calificará más cerca del momento del alta.

Los médicos tendrán entonces la evaluación de riesgos más actualizada para cada paciente y

Les ayudará a elegir a qué pacientes dirigirse para la intervención.

Como parte de la implementación de la solución, el equipo de Respuesta proporcionará personal clínico

desarrollará y proporcionará capacitación.

Además, los resultados pueden pasar por la fase de retroalimentación y el modelo puede probarse a lo largo del tiempo.

monitorear y rastrear a los pacientes que reciben la intervención para que se puedan realizar mejoras dentro

procesos en colaboración con desarrolladores de TI y administradores de bases de datos.

debería desarrollarse.

Este mapa es un ejemplo de una solución implementada a través de una aplicación Cognos.

En este caso, el caso de estudio fue el riesgo de hospitalización de pacientes con diabetes juvenil.

Como el caso de uso de insuficiencia cardíaca congestiva formará la base de esta aplicación.

utilizó la clasificación del árbol de decisión para crear un modelo de riesgo.

El mapa es un análisis interactivo del riesgo previsto por diversas condiciones del paciente y otras características.

Proporciona una visión general del riesgo de hospitalización en todo el país.

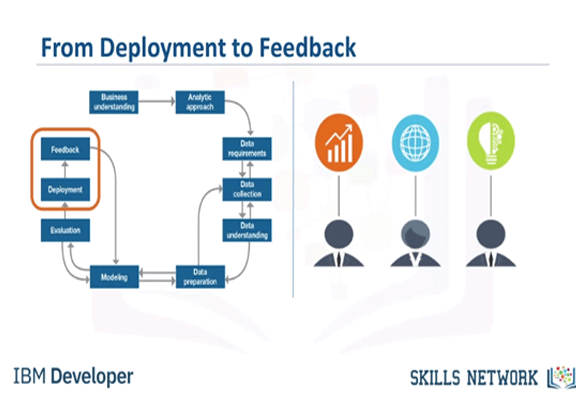
Esta diapositiva es para que los médicos comprendan la combinación de condiciones de este subgrupo de pacientes.

interacción de riesgo por población de pacientes dentro de un nodo determinado del modelo.

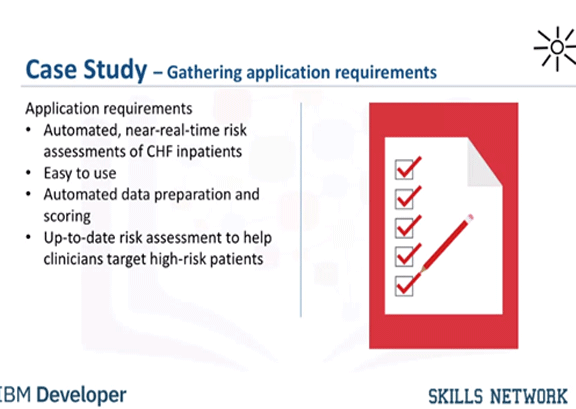
Muestra un informe resumido.

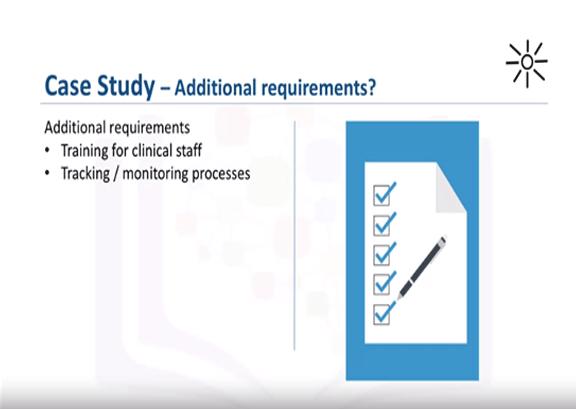
Y este informe, que incluye detalles sobre el riesgo estimado del paciente y su historial clínico.

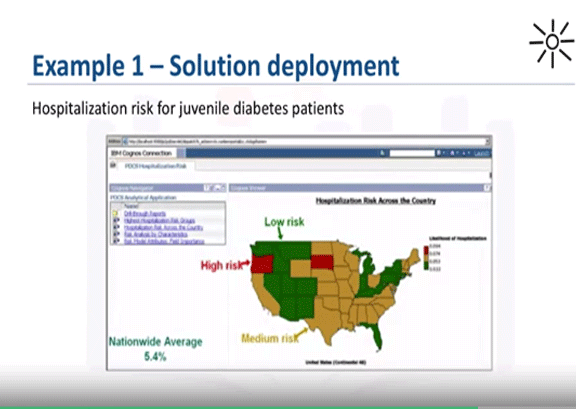
Da un resumen detallado de cada paciente y el médico

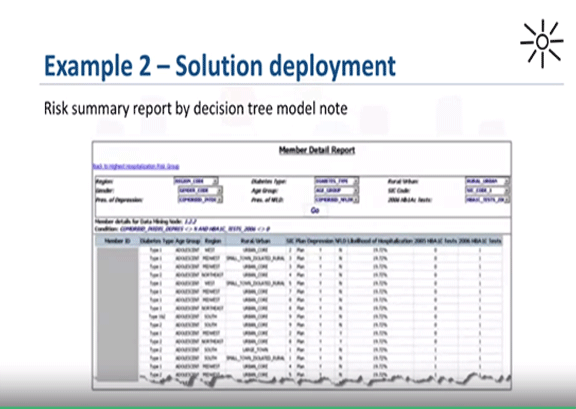
****

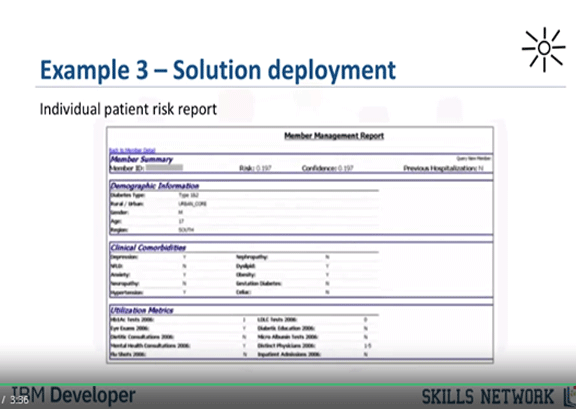
****

****

****

****

****

****

**Comentarios**

El plan para la etapa de retroalimentación incluyó estos pasos:

Primero, se definiría y prepararía el proceso de revisión, con la responsabilidad general

de medir los resultados de un modelo de "vuelo al riesgo" de la población de riesgo por

insuficiencia cardíaca.

Los ejecutivos de manejo clínico serían los responsables del proceso de revisión.

Segundo, se haría seguimiento a los pacientes de insuficiencia cardíaca que recibieran intervención

y se registrarían sus resultados de readmisión.

Tercero, la intervención se mediría para determinar qué tan efectiva era para reducir

las readmisiones.

Por razones éticas, los pacientes de insuficiencia cardíaca no se dividirían en grupos de

control y tratamiento.

En cambio, las tasas de readmisión se compararían antes y después de la implementación del

modelo para medir su impacto.

Luego de las etapas de despliegue y retroalimentación, se revisaría el impacto del programa de intervención

en las tasas de readmisión después del primer año de su implementación.

Entonces se refinaría el modelo, con base en los datos compilados después de la implementación del modelo

y el conocimiento obtenido a lo largo de estas etapas.

Otros refinamientos incluían: Incorporar información de participación

en el programa de intervención, y posiblemente refinar el modelo para incorporar

datos farmacéuticos detallados.

Si recuerdas, la recolección de datos se aplazó inicialmente porque los datos farmacéuticos no

estaban disponibles en ese momento.

Pero después de la retroalimentación y de experimentar el modelo en la práctica, podría decidirse que

agregar esos datos ameritaría la inversión de tiempo y esfuerzo.

También cabe la posibilidad de que otros refinamientos se presentaran por sí mismos

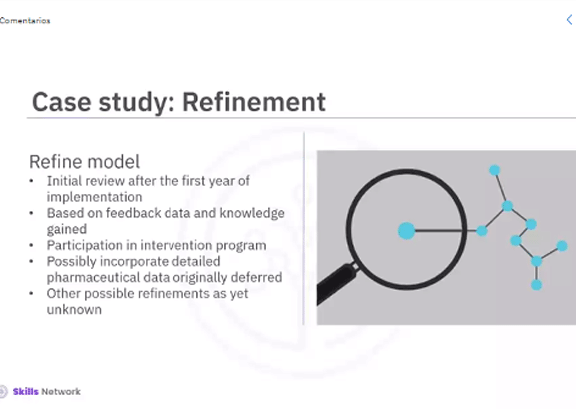
durante la etapa de retroalimentación.

Además, las acciones y procesos de intervención serían revisados y muy seguramente refinados

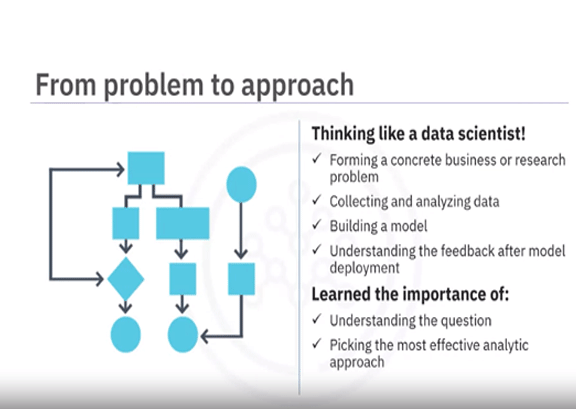
también, con base en la experiencia y el conocimiento adquiridos en el despliegue y la retroalimentación iniciales.

Finalmente, el modelo y las acciones de intervención refinadas se desplegarían de nuevo, manteniendo

el proceso de retroalimentación a lo largo de todo el programa de intervenc

****

**RESUMEN**

****

