

TAXONOMÍA DE LA CIENCIA DE DATOS

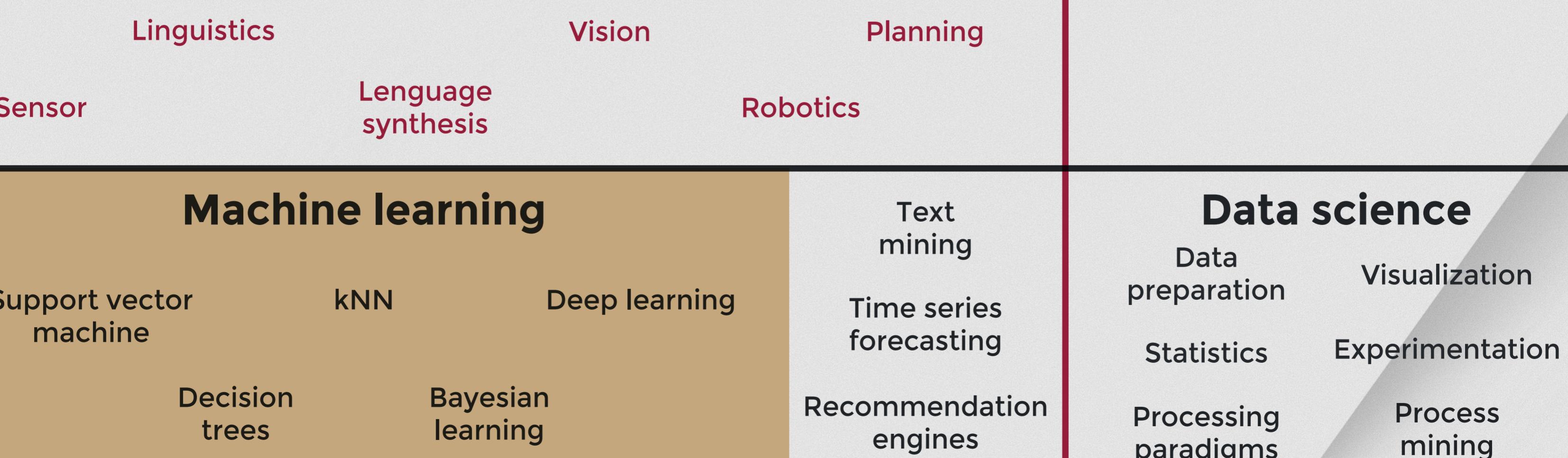
Introducción

- ▶ La ciencia de datos es una colección de técnicas utilizadas para extraer valor de los datos.
- ▶ Las técnicas de ciencia de datos consisten en encontrar patrones, conexiones y relaciones útiles dentro de los datos.
- ▶ La ciencia de datos también se conoce comúnmente como descubrimiento de conocimiento, aprendizaje automático, análisis predictivo y minería de datos. Sin embargo, cada término tiene una connotación ligeramente diferente según el contexto.

Perspectivas analíticas en la ciencia de datos

- 1 La inteligencia artificial, el aprendizaje automático y la ciencia de datos están relacionados entre sí. Como era de esperar, a menudo se usan indistintamente y se fusionan entre sí en medios de comunicación populares y en las empresas.
- 2 Sin embargo, estos tres campos son distintos según el contexto.
- 3 La ciencia de datos comienza con los datos, que pueden ir desde una simple matriz de unas pocas observaciones numéricas hasta una matriz compleja de millones de observaciones con miles de variables.
- 4 La ciencia de datos utiliza ciertos métodos informáticos especializados para descubrir estructuras útiles y significativas dentro de un conjunto de datos.
- 5 En la búsqueda de extraer información útil y relevante de grandes conjuntos de datos, la ciencia de datos toma prestadas técnicas computacionales de las disciplinas de estadística, aprendizaje automático, inteligencia artificial, experimentación y administración de bases de datos.

Artificial intelligence



Inteligencia artificial: la inteligencia artificial se trata de dar a las máquinas la capacidad de imitar el comportamiento humano, en particular las funciones cognitivas. El aprendizaje es una parte importante de la capacidad humana. De hecho, muchos otros organismos vivos pueden aprender.

Algunos ejemplos serían: reconocimiento facial, vehículos automatizados, clasificación del correo según el código postal.

Hay una gran variedad de técnicas que se incluyen en la inteligencia artificial: lingüística, procesamiento del lenguaje natural, ciencia de decisiones, sesgo, visión, robótica, planificación, etc.

Aprendizaje de máquina: el aprendizaje automático puede considerarse un subcampo o una de las herramientas de la inteligencia artificial, ya que proporciona a las máquinas la capacidad de aprender de la experiencia.

La experiencia para las máquinas viene en forma de datos. Los datos que se utilizan para enseñar a las máquinas se denominan datos de entrenamiento. Los algoritmos de aprendizaje automático, también llamados "aprendices", utilizan los datos de entrenamiento para descubrir un modelo o "comportamiento" en función de los datos de entrada y salida del algoritmo.

Por ejemplo, muchas organizaciones, como las plataformas de redes sociales, o los foros utilizan el aprendizaje automático para modelar publicaciones abusivas o fuera de contexto.

Ciencia de datos: la ciencia de datos es un campo interdisciplinario que extrae valor de los datos, y se refiere a la aplicación empresarial del aprendizaje automático, la inteligencia artificial y otros campos cuantitativos como las estadísticas, la visualización y las matemáticas.

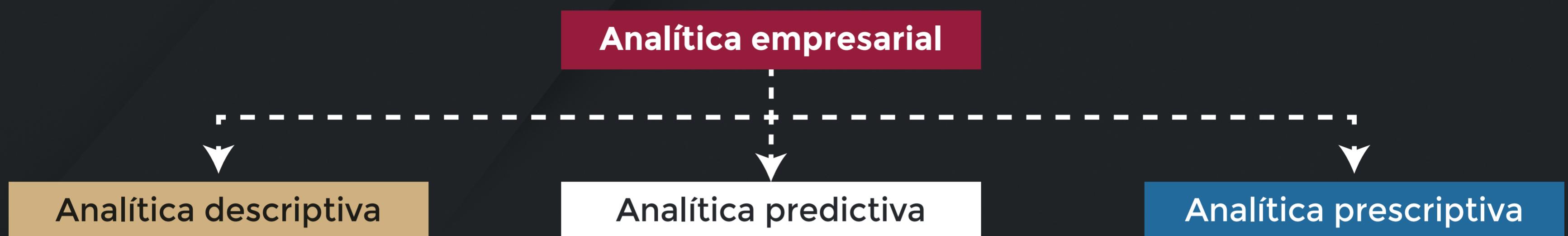
En el contexto de cómo se usa la ciencia de datos hoy en día, se basa en gran medida en el aprendizaje automático y, a veces, se denomina minería de datos.

Ejemplos de casos de ciencia de datos son: motores de recomendación que pueden recomendar películas para un usuario en particular, un modelo de alerta de fraude que detecta transacciones fraudulentas con tarjetas de crédito, predecir clientes que probablemente dejarán una suscripción a página web o pronosticar ingresos para el próximo trimestre en una empresa.



Taxonomía de las metodologías de la ciencia de datos

- ▶ En el campo de la analítica empresarial (business analytics), podemos encontrar tres dominios básicos que clasifican las métodos o técnicas comúnmente utilizadas en la ciencia de datos.



Analítica descriptiva: la analítica descriptiva responde a la pregunta **¿Qué ha sucedido en la organización?**

Es el tipo de análisis más común utilizado por las empresas y generalmente se caracteriza por estadísticas descriptivas, indicadores clave de rendimiento (KPI), tableros u otros tipos de visualizaciones.

El análisis descriptivo resume lo que ha sucedido y que también forma la base de muchos sistemas de alerta de monitoreo continuo, donde las transacciones se comparan con puntos de referencia y se establecen umbrales a partir del análisis de índices y tendencias de datos históricos.

Analítica predictiva: la analítica predictiva es el siguiente paso que se da con la adquisición de conocimiento a partir de la analítica descriptiva y responde a la pregunta **¿Qué podría suceder en la organización?**.

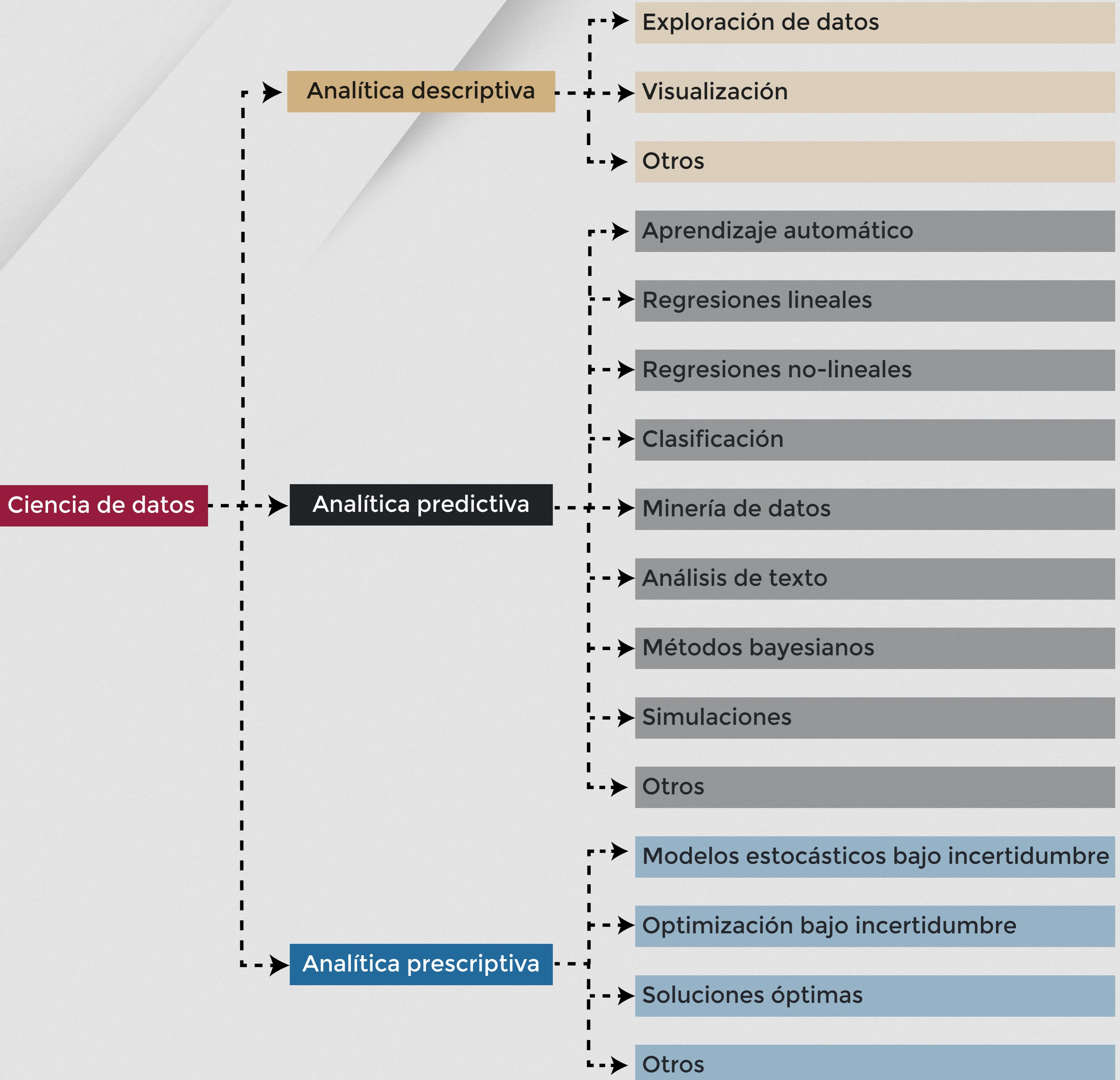
Se caracteriza por modelos predictivos y de probabilidad, pronósticos, análisis estadísticos y modelos de puntuación. Los modelos predictivos utilizan datos históricos acumulados a lo largo del tiempo para realizar cálculos de eventos futuros probables. La mayoría de las empresas utilizan análisis predominantemente descriptivos y recién han comenzado a utilizar análisis predictivos.

Analítica prescriptiva: la analítica prescriptiva responde a la pregunta de **¿Qué se debe hacer en la organización?**

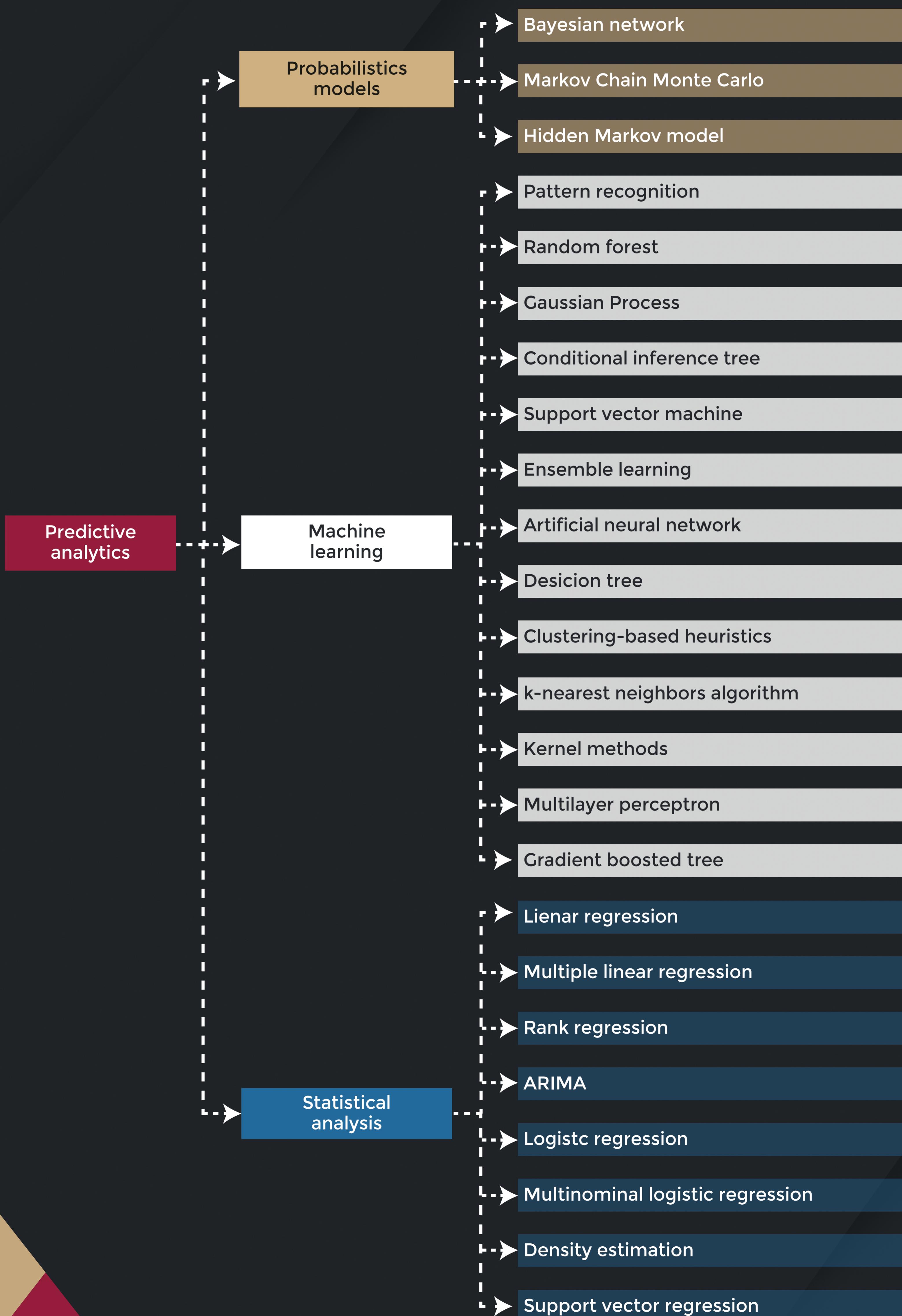
El análisis prescriptivo puede describirse como un enfoque de optimización. El análisis prescriptivo va más allá de lo descriptivo y predictivo al recomendar una o más soluciones y mostrar el resultado probable de cada una.



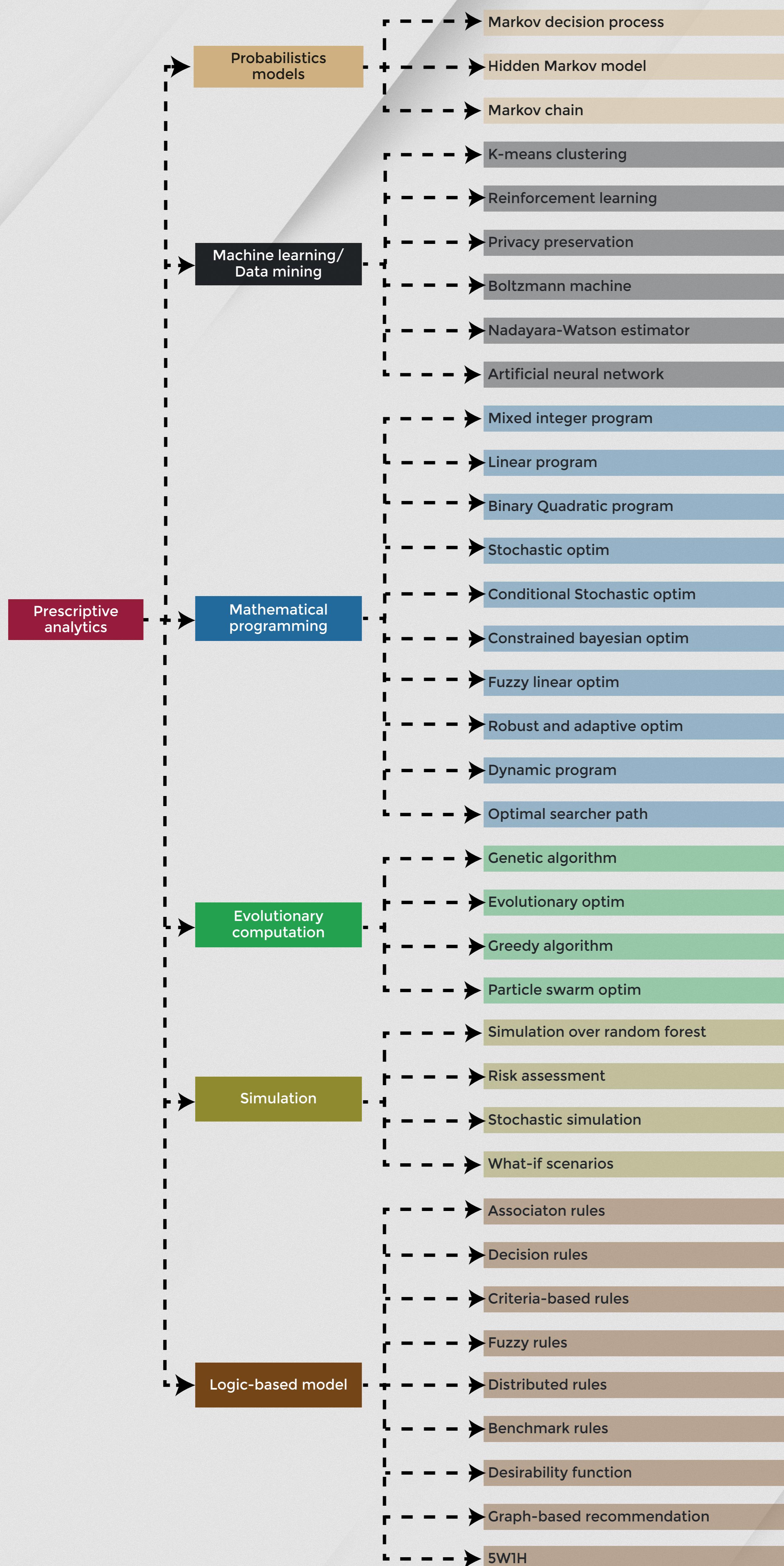
Clasificación de los métodos de la analítica empresarial



Clasificación de los métodos para la analítica predictiva



Clasificación de los métodos para la analítica prescriptiva



Comparación de algoritmos

De ciencia de datos

- ▶ Los problemas de la ciencia de datos también se pueden clasificar en tareas como: clasificación, regresión, análisis de asociación, agrupamiento, detección de anomalías, motores de recomendación, selección de características, pronóstico de series temporales, minería de texto y aprendizaje profundo.
- ▶ En la tabla siguiente, se presenta una comparación de los algoritmos utilizados en la ciencia de datos.

Clasificación: predecir unavariante-objetivo categórica

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Árboles de decisión	Particiona los datos en subconjuntos más pequeños donde cada subconjunto contiene (principalmente) respuestas de una clase (ya sea "sí" o "no").	Reglas para particionar un conjunto de datos basado en los valores del diferente predictores.	Sin restricciones en el tipo de variables para predictores.	La etiqueta no puede ser numérica. Debe ser categórica.	Intuitivo para explicar a usuarios empresariales no técnicos. La normalización de predictores no es necesario.	Puede existir un sobreajuste de los datos. Pequeños cambios en la entrada de datos pueden producir sustancialmente árboles diferentes. Seleccionarlos parámetros de entregada puede ser desafiante.	Marketing segmentación, detección de fraude.
Inducción de reglas	Modela la relación entre entrada y salida a través de la deducción de reglas "IF/THEN" del conjunto de datos.	Un conjunto de reglas organizadas que contienen un antecedente (input) y consecuencias (output class).	Sin restricciones. Acepta entradas categóricas, numéricas, y binarias.	Predicción de la variable objetivo, la cual es categórica.	El modelo puede ser fácilmente explicado a usuarios empresariales. Fácil de desplegar en casi cualquier herramienta y aplicación.	Divide el conjunto de datos en forma rectilínea.	Fabricación, aplicaciones dónde la descripción del modelo es necesario.
k-vecinos más próximos (k-nearest neighbors)	Cualquier nuevo punto de dato desconocido se compara con puntos de datos conocidos similares en el conjunto de entrenamiento.	El modelo es el conjunto de datos completo de entrenamiento.	Sin restricciones. Sin embargo, el cálculo de distancias funciona mejor con datos numérico. Los datos deben ser normalizados.	Predicción de la variable objetivo, la cual es categórica.	Requiere muy poco tiempo para construir el modelo. Maneja correctamente los atributos que faltan en el registro desconocido. Funciona con relaciones no lineales.	El tiempo de ejecución de la implementación y los requisitos de almacenamiento serán costosos. Selección arbitraria de valor de k. Sin descripción del modelo.	Procesamiento de imágenes, aplicaciones en las que es aceptable un tiempo de respuesta más lento.
Clasificador bayesiano (Naive Bayesian)	Predice la clase de salida según el teorema de Bayes mediante el cálculo de la probabilidad condicional de clase y la probabilidad previa.	Una tabla de búsqueda de probabilidades y probabilidades condicionales para cada atributo con una clase de salida.	Sin restricciones. Sin embargo, el cálculo de probabilidad funciona mejor con atributos categóricos.	Predicción de probabilidad para todos los valores de clase, junto con la clase ganadora.	Tiempo requerido para modelo y desplegar es mínimo. Gran algoritmo para evaluación comparativa. Fuerte estadístico base.	Capacitación datos colocar necesita ser representante muestra de población y necesita tener completo combinaciones de aporte y producción. Atributos necesidad a ser independiente.	Correo basura detecciones, texto minería.
Redes neuronales artificiales	Modelo computacional y matemático inspirado en el sistema nervioso biológico. Los pesos en la red aprender a reducir el error entre lo actual y la predicción.	Una red topología de capas y pesos para procesar datos de entrada.	Todos los atributos deben ser numéricos.	Predicción de la variable objetivo (etiqueta), la cual es categórica.	Es bueno para modelar relaciones no lineales. Tiempo de respuesta rápido en la implementación.	No hay manera fácil de explicar el funcionamiento interno del modelo. Requiere datos de preprocessamiento. No se pueden manejar atributos faltantes.	Reconocimiento de imágenes, detección de fraudes, aplicaciones con tiempo de respuesta rápida.
Máquinas de vectores de soporte (support vector machines)	Algoritmo de detección de límites que identifica/ define límites multidimensionales separando puntos de datos que pertenecen a diferentes clases.	El modelo es una ecuación de vectores que permite clasificar nuevos puntos de datos en diferentes regiones (clases)..	Todos los atributos deben ser numéricos.	Predicción de la variable objetivo (etiqueta), la cual es categórica o numérica. Es bueno para modelar relaciones no lineales.	Extremadamente robusto contra el sobreajuste. Pequeños cambios en los datos de entrada no afecta el límite y, por lo tanto, no ofrece diferentes resultados. Es bueno para modelar relaciones no lineales.	Desempeño computacional durante la fase de entrenamiento puede ser lento. Esto puede ser atribuido al esfuerzo requerido para optimizar combinación de parámetros.	Reconocimiento, óptico de caracteres, detección de, modelado de ventos categorizados como "cisne negro".

Clasificación: predecir una variable-objetivo categórica

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Modelos en conjunto (ensemble models)	Emplea un numero independiente de modelos para hacer una predicción y agrega la predicción final.	Un meta-modelo con modelos base individual y un agregador.	Superconjunto de restricciones desde el modelo base utilizado.	Predicción para todos los valores de clase con una clase ganadora .	Reduce el error de generalización. Toma en consideración diferentes espacios de búsqueda.	Lograr la independencia del modelo es complicado. Difícil de explicar el funcionamiento interno del modelo.	La mayoría de los clasificadores prácticos son conjunto (ensemble).

Regresión: predecir una variable-objetivo numérica

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Regresión Lineal	Clásico modelo predictivo que expresa la relación entre parámetros de entradas y salidas forma de una ecuación.	El modelo consta de coeficientes para cada predictor de entrada y su significancia estadística.	Todos los atributos deben ser numéricos.	La etiqueta puede ser numérica o binomial.	El caballo de batalla de la mayoría de los modelados predictivos. Fácil de usar y explicar a usuarios comerciales no técnicos.	No puede manejar datos faltantes. Datos categóricos no son directamente utilizables, pero requieren transformación a numérico.	Cualquier situación que requiere predecir un valor numérico continuo.
Regresión logística	Técnicamente, es un método de clasificación. Pero estructuralmente es similar a la regresión lineal.	El modelo consiste de coeficientes para cada predictor de entrada que se relaciona con el "logit". Transformando el logit en probabilidades de ocurrencia (de cada clase) completa el modelo.	Todos los atributos deben ser numéricos.	La etiqueta puede ser binomial.	Uno de los métodos más comunes para clasificación. Computacionalmente eficiente.	No puede manejar datos faltantes. No es intuitivo cuando se trata con un gran número de predictores.	Marketing, cualquier problema general de dos clases .

Análisis de asociación: proceso no supervisado para encontrar relaciones entre elementos

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Crecimiento de patrón frecuente (FP-growth) y Apriori	Mide la fortaleza de coocurrencia entre un elemento con otro.	Encuentra reglas simples, fáciles de entender como {Manzanas, Leche} -> {Cerveza}.	Formato de transacciones con elementos en las columnas y transacciones en las filas.	Lista de reglas importantes desarrolladas desde el conjunto datos.	Enfoque no supervisado con entradas mínimas de usuario. Fácil de entender las reglas.	Requiere preprocessamiento si la entrada es de diferente formato.	Motores de recomendación, venta cruzada, y sugerencias de contenido.



Agrupación: proceso no supervisado para encontrar grupos significativos en los datos

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
k - Medios	El conjunto de datos es dividido en clústeres (grupos) k al encontrar k centroides.	El algoritmo encuentra k centroides y todos los puntos de datos son asignados hacia el centroide más cercano, el cual forma un clúster o grupo.	Sin restricciones. Sin embargo, el cálculo de la distancia funciona mejor con datos numéricos. Los datos deben ser normalizados.	El conjunto de datos es anexado por uno de las etiquetas de grupo k.	Fácil de implementar. Puede ser usado para reducción dimensional.	Especificación de k es arbitrario y puede no encontrar clústeres naturales. Sensible a valores atípicos.	Segmentación de clientes, detección de anomalías, aplicaciones dónde agrupamiento globular es natural.
DBSCAN	Identifica clústeres como una área de alta densidad rodeado por áreas de baja densidad.	Lista de clústeres y puntos de datos asignados. El clúster por defecto 0 contiene datos de ruido.	Sin restricciones. Sin embargo, el cálculo de la distancia funciona mejor con datos numéricos. Los datos deben ser normalizados.	Las etiquetas de clúster basadas en clústeres identificados.	Encuentra los clústeres naturales de cualquier forma. No hay necesidad de mencionar número clústeres.	Especificación de densidad de parámetros A puente entre dos clústeres puede fusionar el grupo. No puede agrupar variando la densidad de los puntos de datos.	Aplicaciones dónde los clústeres son formas no globulares y cuando el numero anterior de agrupaciones naturales son desconocidas.
Mapas auto organizados	Una técnica de agrupamiento visual con bases en redes neurales y agrupamiento prototípico.	Red bidimensional donde los puntos de datos similares se organizan uno junto al otro.	Sin restricciones. Sin embargo, el cálculo de la distancia funciona mejor con datos numéricos. Los datos deben ser normalizados.	No se identifican clústeres explícitos. Los puntos de datos similares ocupan la misma celda o se colocan uno al lado del otro en la vecindad.	Una forma visual de explicar los clústeres. Reduce los datos multidimensionales a dos dimensiones.	El usuario especifica el número de centroides (topología). No encuentra clústeres naturales en los datos.	Diversas aplicaciones incluyendo exploración de datos visual, sugerencias de contenido, y reducción dimensional.

Detección de anomalías: técnicas supervisadas y no supervisadas para encontrar valores atípicos en los datos

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Basado en la distancia	Valor atípico identificado en función de la distancia al k-ésimo vecino más cercano.	Todos los puntos de datos son asignados un puntaje de distancia basado en el vecino más cercano.	Acepta atributos numéricos y categóricos. Normalización es requerida debido a que la distancia es calculada.	Cada punto de dato tiene un puntaje de distancia. Cuanto mayor sea la distancia, más probable es que el punto de dato sea un valor atípico.	Fácil de implementar. Funciona bien con atributos numéricos.	Especificación de k es arbitrario.	Detección de fraude, técnica de preprocesamiento.
Basado en la densidad	El valor atípico se identifica en función de puntos de datos en regiones de baja densidad.	Todos los puntos de datos son asignados un puntaje de densidad basado en el vecino.	Acepta atributos numéricos y categóricos. Normalización es requerida debido a que la distancia es calculada.	Cada punto de dato tiene un puntaje de densidad. Cuanto mayor sea la densidad, más probable es que el punto de dato sea un valor atípico.	Fácil de implementar. Funciona bien con atributos numéricos.	Especificación del parámetro de distancia por el usuario. Incapacidad para identificar regiones de densidad variable.	Detección de fraude, técnica de preprocesamiento.
Factor de valor atípico local	El valor atípico se identifica con base en el cálculo de la densidad relativa en los vecindarios.	Todos los puntos de datos son asignados un puntaje de densidad relativa basado en el vecino.	Acepta atributos numéricos y categóricos. Normalización es requerida debido a que la densidad es calculada.	Cada punto de datos tiene un puntaje de densidad. Cuanto menor sea la densidad relativa, más probable es que el punto de dato sea un valor atípico.	Puede manejar la variación de densidad.	Especificación del parámetro de distancia por el usuario.	Detección de fraude, técnica de preprocesamiento.

Sistema de recomendación: encontrar la preferencia del usuario de un artículo

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Filtrado colaborativo - basado en el vecino	Encuentre una cohorte de usuarios que proporcionaron calificaciones similares. Obtiene la calificación de resultado de los usuarios de la cohorte.	Usuarios similares o artículos tienen gustos similares.	Matriz de calificaciones con preferencias de usuario-artículo.	Matriz completa de calificaciones.	La única entrada necesaria es la matriz de calificaciones Agnóstico de dominio.	Problemas de arranque en frío (cold start) para nuevos usuarios y artículos. La computación crece linealmente con la cantidad de artículos y usuarios.	Comercio electrónico, música, nueva conexión de recomendaciones.
Filtrado colaborativo - factorización de matriz latente	Descompone la matriz usuario-artículo en dos matrices (P y Q) con factores latentes. Rellene los valores en blanco en la matriz de calificaciones por el producto punto de P y Q.	La preferencia del usuario por un elemento se puede explicar mejor por su preferencia por el carácter de un artículo (inferido).	Matriz de calificaciones con preferencias de usuario-artículo.	Matriz completa de calificaciones.	Funciona en matriz dispersa. Más preciso que el filtrado colaborativo basado en el vecindario.	No puedo explicar por qué se hace la predicción.	Recomendaciones de contenido.
Filtrado basado en contenido	Abstactae las características del artículo y crear el perfil del artículo. Usa el perfil del artículo para evaluar la preferencia del usuario por los atributos del perfil del artículo.	Recomienda artículos similares a los que le gustaron al usuario en el pasado.	Matriz de calificación de artículo-usuario y perfil de artículo.	Matriz completa de calificaciones.	Soluciona el problema de arranque en frío (cold start) para artículos nuevos. Puede proporcionar explicaciones sobre por qué se hace la recomendación.	Requiere conjunto de datos de perfil de artículo. Los recomendaciones son específicos del dominio.	Recomendación de música de Pandora y la indexación de citas de CiteSeer.
Basado en contenido - Modelos de aprendizaje supervisado	Un modelo de clasificación o regresión personalizado para cada usuario del sistema. Aprende un clasificador basado en los gustos o disgustos de los usuarios de un artículo y su relación con los atributos del artículo.	Cada vez que un usuario prefiere un artículo, es un voto de preferencia por los atributos del artículo.	Matriz de calificación de artículo-usuario y perfil de artículo.	Matriz completa de calificaciones.	Cada usuario tiene un modelo separado y se puede personalizar de forma independiente. Hiper personalización.	Almacenamiento y tiempo computacional.	Contenido de comercio electrónico, y conexión de recomendaciones.

Predicción de series temporales: predicción del valor futuro de una variable

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Descomposición	Descompone la serie de tiempo en tendencia, estacionalidad, y ruido. Pronostica los componentes.	Modela para los componentes individuales.	Valores históricos.	Valor pronosticado.	Aumenta comprensión de la serie de tiempo al visualizar los componentes.	La precisión depende de los modelos utilizados para los componentes.	Aplicaciones dónde la explicación de componentes es importante.
Suavizado exponencial	El valor futuro es una función de observaciones pasadas.	Aprende los parámetros de la ecuación de suavizado de datos históricos.	Valores históricos.	Valor pronosticado.	Se aplica a una amplia gama de series temporales con o sin tendencia o estacionalidad.	La estacionalidad múltiple en los datos hace que los modelos sean engorrosos.	Casos dónde tendencia o estacionalidad no es evidente.
ARIMA	El valor futuro es la función de puntos de datos pasados correlacionados automáticamente y el promedio móvil de las predicciones.	Parámetro para los valores (p,d,q), AR, y coeficientes AR y.	Valores históricos.	Valor pronosticado.	Forma una línea base estadística para exactitud del modelo.	Los valores óptimos p, d, q valor son desconocidos.	Se aplica en casi todo tipos de datos de series de tiempo.

Predicción de series temporales: predicción del valor futuro de una variable

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Windowing-basado en aprendizaje de máquina	Crear un conjunto de datos transversales con entradas de tiempo retrasadas.	Modelos de aprendizaje de máquina como regresión, red neural redes, etc.	Valores históricos.	Valor pronosticado.	Usa cualquier enfoque de aprendizaje de máquina en datos seccionales.	El tamaño de las ventanas, horizonte, y valores saltantes son arbitrarios.	Aplica a casos de usuario donde la serie de tiempo tiene tendencia y/o estacionalidad.

Selección de características: selección de los atributos más importantes

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
PCA (análisis de componentes principales) basado en filtros	Combina los atributos más importantes en un número menor de atributos transformados.	Cada componente principal es una función de atributos del conjunto de datos.	Atributos numéricos.	Atributos numéricos (set reducido). No necesariamente requieren una etiqueta.	Manera eficiente para extraer predictores que no están correlacionados entre sí. Ayuda a aplicar el principio de Pareto en la identificación de atributos con más alta varianza.	Sensible a efectos de escala, es decir, requiere la normalización de los valores de los atributos antes de su utilización. Centrarse en la varianza a veces da como resultado la selección de atributos ruidosos.	La mayoría de los conjuntos de datos con valores numéricos requieren una reducción de dimensiones.
Ganancia de información (basado en filtros)	Seleccionan atributos basados en relevancia para la variable objetivo o etiqueta.	Similar a modelo de árbol de decisión.	Sin restricciones en el tipo de variable para los predictores.	Conjuntos de datos requieren una etiqueta. Solo puede ser aplicado en conjuntos de datos con etiqueta nominal.	Similar a árboles de decisión.	Similar a árboles de decisión.	Aplicaciones para selección de características donde la variable objetivo es categórica o numérica.
Chi-cuadrado (basado en filtros)	Seleccionan atributos basados en relevancia para la variable objetivo o etiqueta.	Usa la prueba de chi-cuadrado de independencia para relacionar predictores de etiqueta.	Categórico (atributos polinomiales).	Conjuntos de datos requieren una etiqueta. Solo puede ser aplicado en conjuntos de datos con etiqueta nominal.	Extremadamente robusto. A rápido y eficiente esquema para identificar que variable categórica a seleccionar para un modelo predictivo.	A veces difícil de interpretar.	Aplicaciones para selección de características donde la variable objetivo es categórica.
Selección hacia adelante (Forward selection-wrap per based)	Seleccionan atributos basados en relevancia para la variable objetivo o etiqueta.	Trabaja en conjunción con métodos de modelado tales como regresión.	Todos los atributos deben ser numéricos.	La etiqueta puede ser numérica o binomial.	Problemas de multicolinealidad pueden ser evitados. Acelera la fase de entrenamiento del proceso de modelado.	Una vez que una variable es añadida al conjunto, nunca es retirada en iteraciones subsecuentes incluso si su influencia en la variable objetivo disminuye.	Conjuntos de datos con un gran número de variables de entrada dónde la selección de características es requerido.
Eliminación hacia atrás (Backward elimination wrapper-based)	Seleccionan atributos basados en relevancia para la variable objetivo o etiqueta.	Trabaja en conjunción con métodos de modelado tales como regresión.	Todos los atributos deben ser numéricos.	La etiqueta puede ser numérica o binomial.	Problemas de multicolinealidad pueden ser evitados. Acelera la fase de entrenamiento del proceso de modelado.	Es necesario empezar modelo completo, el cual puede ser computacionalmente intensivo.	Conjuntos de datos con pocas variables de entrada dónde la selección de características es requerido.

Aprendizaje profundo: entrenamiento usando múltiples capas de representación de datos

Algoritmo	Descripción	Modelo	Input	Output	Pros	Cons	Casos
Convolucional	Basado en el concepto de aplicar filtros a la representación bidimensional de datos entrantes, como imágenes. El aprendizaje automático se utiliza para determinar automáticamente los pesos correctos para los filtros.	Un tensor de típicamente tres o más dimensiones. Dos de las dimensiones corresponden a la imagen, mientras que una tercera se usa a veces para la codificación de color/canal.	Por lo general, la salida de la capa convolucional se aplana y pasa a través de una capa densa o completamente conectada que generalmente termina en una capa de salida softmax.	Red muy potente y de propósito general. El número de pesos a aprender en la capa de conv no es muy alto.	Usa cualquier enfoque de aprendizaje de máquina en datos seccionales.	Para la mayoría de los problemas prácticos de clasificación, las capas conv deben combinarse con capas densas, lo que da como resultado una gran cantidad de pesos para entrenar y, por lo tanto, pierde las ventajas de velocidad de una capa conv pura.	Clasifica casi cualquier dato donde la información espacial esté altamente correlacionada, como imágenes. Incluso los datos de audio pueden convertirse en imágenes (usando transformadas de Fourier) y clasificarse a través de redes de conv.
Recurrente	Así como las redes conv están especializadas para analizar datos correlacionados espacialmente, las redes recurrentes están especializadas para datos correlacionados temporalmente: secuencias. Los datos pueden ser secuencias de números, señales de audio o incluso imágenes.	Una secuencia de cualquier tipo (serie temporal, texto, discurso, etc).	RNN puede procesar secuencias y produce otras secuencias (muchas a muchos), o produce un tensor fijo (muchos a uno).	RNN puede procesar secuencias y produce otras secuencias (muchas a muchos), o produce un tensor fijo (muchos a uno).	A diferencia de otros tipos de redes neuronales, las RNN no tienen restricciones de que la forma de entrada de los datos sea de dimensión fija.	Los RNN sufren gradientes que desaparecen (o explotan) cuando las secuencias son muy largas. Los RNN tampoco son aptos para muchas capas apiladas por la misma razón.	Pronóstico de series de tiempo, situaciones de procesamiento de lenguaje natural como traducción automática, subtítulos de imágenes.

Referencias:

- ▶ Kotu, V., & Deshpande, B. (2018). *Data science: concepts and practice*. Morgan Kaufmann.
- ▶ Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems*, 25, 29-44.
- ▶ Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Construing the big data based on taxonomy, analytics and approaches. *Iran Journal of Computer Science*, 1(4), 237-259.
- ▶ Raghupathi, W., & Raghupathi, V. (2021). Contemporary Business Analytics: An Overview. *Data*, 6(8), 86.
- ▶ Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50, 57-70