

PREGUNTAS DE DATOS MASIVOS PARA ESTUDIAR (POWER 2)

PREGUNTA IMPORTANTE POWER 1: ¿QUÉ SIGNIFICA QUE SPARK TIENE UN MOTOR DE EJECUCIÓN FÍSICA?

1. ¿QUÉ SON LOS SISTEMAS DE RECOMENDACIÓN? ¿CÓMO TRABAJAN?
2. ¿CÓMO SE GENERAN LOS DATOS DE LOS SISTEMAS DE RECOMENDACIÓN? (¿QUÉ SIGNIFICA QUE LOS USUARIOS SE ENCUENTRAN EN UN CICLO DE RETROALIMENTACIÓN?)
3. ¿CUÁLES SON LOS ELEMENTOS DE LOS SISTEMAS DE RECOMENDACIÓN?
4. ¿CUÁL ES EL PROCESO DE LOS SISTEMAS DE RECOMENDACIONES?
5. NOMBRA ALGÚN USO COMÚN DE LOS SISTEMAS DE RECOMENDACIÓN.
6. ¿CUÁL ES EL OBJETIVO DE LOS SISTEMAS DE RECOMENDACIÓN?
7. ¿DE QUÉ SE BASA LA ARQUITECTURA DE LOS SISTEMAS DE RECOMENDACIÓN?
8. ¿QUÉ TIPO DE ALGORITMOS O RELACIONES SE PUEDEN UTILIZAR EN SISTEMAS DE RECOMENDACIÓN?
9. ¿QUÉ TIPOS DE INFORMACIÓN HAY PARA LOS SISTEMAS DE RECOMENDACIÓN?
9. ¿QUÉ TIPOS DE INFORMACIÓN HAY PARA LOS SISTEMAS DE RECOMENDACIÓN?
10. ¿CUÁLES SON LAS TÉCNICAS MÁS USADAS?
11. ¿QUÉ MODELOS DE RECOMENDACIÓN HAY?
12. ¿QUÉ SISTEMAS DE CALIFICACIÓN USUARIO/ OBJETO HAY?
13. ¿CUÁLES SON LOS RETOS DE LOS SISTEMAS DE RECOMENDACIÓN?
14. ¿CUÁLES SON LOS PASOS PARA EL PROCESAMIENTO DE DATOS CON MACHINE LEARNING?
15. ¿CUÁLES SON LOS CONCEPTOS DE APACHE SPARK PARA EL PROCESAMIENTO CON MACHINE LEARNING?
16. ¿QUÉ ES EL APRENDIZAJE SUPERVISADO? EXPLICA UN PROCESO MACHINE LEARNING CON APRENDIZAJE SUPERVISADO CON SPARK Y DA UN POSIBLE CASO DE USO.
17. EXPLICA EL PROCESO DE ENTRENAMIENTO DE UN MODELO DE MACHINE LEARNING, EXPLICA CON QUÉ EMPIEZA EL PROCESO Y CON QUÉ FINALIZA. ¿POR QUÉ ES TAN IMPORTANTE EL PROCESO DE ENTRENAMIENTO?
18. EXPLICA QUÉ ES UN MODELO O PROBLEMA DE REGRESIÓN (QUÉ OBJETIVO PREDICTIVO TIENE)
19. EXPLICA QUÉ ES UN MODELO O PROBLEMA DE CLASIFICACIÓN (QUÉ OBJETIVO PREDICTIVO TIENE)
20. NOMBRA CINCO ALGORITMOS DIFERENTES DISPONIBLES EN SPARK MLIB Y QUÉ FUNCIONALIDAD APORTAN (CLASIFICACIÓN O REGRESIÓN)

21. ¿QUÉ DATOS DE ENTRADA RECIBE UN SISTEMA DE RECOMENDACIÓN? ¿QUÉ DATOS DE SALIDA CALCULA EL SISTEMA DE RECOMENDACIÓN?

22. EXPLICA PASO A PASO EL FLUJO DE TRABAJO DE EL PROCESAMIENTO DE DATOS CON MACHINE LEARNING Y QUÉ ESTRUCTURAS INTERACTÚAN EN EL FLUJO EN SPARK.

24. EXPLICA LA ESTRUCTURA: TRANSFORMADORES. PON UN EJEMPLO DE TRANSFORMADOR.

25. EXPLICA LA ESTRUCTURA: ESTIMADORES. PON UN EJEMPLO DE ESTIMADOR.

26. EXPLICA LA ESTRUCTURA: EVALUADORES. PON UN EJEMPLO DE EVALUADOR

27. EXPLICA LA ESTRUCTURA: PIPELINES.

PREGUNTA IMPORTANTE POWER 1: ¿QUÉ SIGNIFICA QUE SPARK TIENE UN MOTOR DE EJECUCIÓN FÍSICA?

SIGNIFICA QUE SPARK ES CAPAZ DE EJECUTAR TAREAS DE PROCESAMIENTO SOBRE UN CLUSTER GRACIAS A SU MOTOR, QUE TRANSFORMA DIRECTRICES LÓGICAS A NIVEL DE USUARIO EN DIRECTRICES FÍSICAS (MÁS CONCRETAS) SOBRE CÓMO OBTENER EL RESULTADO QUE ESPERA EL USUARIO.

1. ¿QUÉ SON LOS SISTEMAS DE RECOMENDACIÓN? ¿CÓMO TRABAJAN?

SON HERRAMIENTAS QUE OFRECEN PRODUCTOS, SERVICIOS O CONTENIDO A LOS USUARIOS EN BASE A SUS COMPORTAMIENTOS ANTERIORES. UTILIZAN UN REGISTRO HISTÓRICO DE LAS ACCIONES O PREFERENCIAS DE LOS USUARIOS PARA PREDECIR RECOMENDACIONES QUE PUEDEN INTERESAR AL USUARIO.

2. ¿CÓMO SE GENERAN LOS DATOS DE LOS SISTEMAS DE RECOMENDACIÓN? (¿QUÉ SIGNIFICA QUE LOS USUARIOS SE ENCUENTRAN EN UN CICLO DE RETROALIMENTACIÓN?)

LOS USUARIOS NO SOLO RECIBEN RECOMENDACIONES, SINO QUE SE ENCUENTRAN EN UN CICLO DE RETROALIMENTACIÓN QUE QUIERE DECIR QUE SUS ACCIONES O PREFERENCIAS SE UTILIZARÁN PARA FUTURAS RECOMENDACIONES A ÉL MISMO O A OTROS USUARIOS SIMILARES Y DE ESTA FORMA SE CREA UN REGISTRO HISTÓRICO PARA LOS SISTEMAS DE RECOMENDACIÓN.

3. ¿CUÁLES SON LOS ELEMENTOS DE LOS SISTEMAS DE RECOMENDACIÓN?

LOS ELEMENTOS DE LOS SISTEMAS DE RECOMENDACIÓN SON TRES: LOS ALGORITMOS DE MACHINE LEARNING QUE SON MÉTODOS MATEMÁTICOS Y ESTADÍSTICOS QUE PERMITEN APRENDER DE LOS DATOS Y HACER PREDICIONES, LOS MODELO DE MACHINE LEARNING QUE ES EL RESULTADO DE ENTRENAR EL ALGORITMO DE MACHINE LEARNING CON PARTE DE NUESTROS DATOS Y LOS RESULTADOS QUE SE CONSIGUEN GRACIAS A LA APLICACIÓN DEL MODELO A NUEVOS DATOS.

4. ¿CUÁL ES EL PROCESO DE LOS SISTEMAS DE RECOMENDACIONES?

PRIMERO SE CARGAN LOS DATOS, DESPUÉS SE PROCESAN LOS DATOS, ES DECIR, SE LIMPIAN Y SE TRANSFORMAN. A CONTINUACIÓN ENTRA LA CIENCIA DE DATOS QUE APLICA MÉTODOS DE MACHINE LEARNING Y ESTADÍSTICA PARA CONSTRUIR EL MODELO, POR ÚLTIMO ENTRA LA INGENIERÍA DE DATOS QUE DESARROLLA Y MANTIENE LA INFRAESTRUCTURA PARA EL ALMACENAMIENTO.

5. NOMBRA ALGÚN USO COMÚN DE LOS SISTEMAS DE RECOMENDACIÓN.

PARA EL E-COMMERCE, PARA LAS EMPRESAS DE MARKETING COMO POR EJEMPLO EL ANÁLISIS DE GESTAS VACÍAS O DE CAMPAÑAS DE EMAIL, O PARA LA BÚSQUEDA DE CONTENIDO EN PLATAFORMAS COMO NETELIX O SPOTIFY E INCLUSO PARA LA PROTECCIÓN DE LA CYBER DEFENSA.

6. ¿CUÁL ES EL OBJETIVO DE LOS SISTEMAS DE RECOMENDACIÓN?

EL OBJETIVO ES AYUDAR A LOS USUARIOS A DESCUBRIR INFORMACIÓN RELEVANTE PARA ELLOS COMO UN ARTÍCULO PARA VENDER, UNA ACCIÓN PARA HACER O UNA NOTIFICACIÓN PARA LEER.

7. ¿DE QUÉ SE BASA LA ARQUITECTURA DE LOS SISTEMAS DE RECOMENDACIÓN?

LA ARQUITECTURA DE LOS SISTEMAS DE RECOMENDACIÓN SE BASA EN:

LOS USUARIOS, QUE ALMACENAN TODA ESA INFORMACIÓN SOBRE ACCIONES, PREFERENCIAS...

LOS OBJETOS, QUE ALMACENAN INFORMACIÓN SOBRE LOS ARTÍCULOS, PRODUCTOS....

LOS CALIFICADORES, QUE EVALÚAN NUMÉRICAMENTE LA INTERACCIÓN USUARIO–OBJETO

Y LOS MODELOS, QUE SE ENCARGAN DE PREDECIR RECOMENDACIONES PARA LOS USUARIOS EN FUNCIÓN DE LOS USUARIOS Y LOS CALIFICADORES.

8. ¿QUÉ TIPO DE ALGORITMOS O RELACIONES SE PUEDEN UTILIZAR EN SISTEMAS DE RECOMENDACIÓN?

EL PRIMER ALGORITMO O RELACIÓN ES USUARIO–ITEM QUE SE CENTRA EN LA SIMILITUD ENTRE USUARIOS E ÍTEMS COMO POR EJEMPLO UN USUARIO QUE COMPRA UN ELEMENTO O UN USUARIO QUE REPRODUCE UNA CANCIÓN. EL SEGUNDO ALGORITMO ES ITEM–ITEM QUE SE CENTRA EN LA SIMILITUD ENTRE OBJETOS POR EJEMPLO DOS CANCIONES DEL MISMO GÉNERO O TINTA COMPATIBLE PARA UNA IMPRESORA. LA ÚLTIMA RELACIÓN ES USUARIO–USUARIO QUE SE CENTRA EN LA SIMILITUD ENTRE DOS USUARIOS COMO PUEDE SER EL MISMO IDIOMA O CONEXIONES EN COMÚN EN REDES SOCIALES.

9. ¿QUÉ TIPOS DE INFORMACIÓN HAY PARA LOS SISTEMAS DE RECOMENDACIÓN?

ENTRE LOS TIPOS DE INFORMACIÓN ESTÁN LOS COMPORTAMIENTOS DEL USUARIO, QUE GUARDAN INFORMACIÓN DEL HISTORIAL DEL USUARIO. LA DEMOGRAFÍA DEL USUARIO, QUE GUARDA INFORMACIÓN PERSONAL DEL USUARIO COMO LA UBICACIÓN, LA EDAD O EL GÉNERO Y POR ÚLTIMO, LOS ATRIBUTOS DE LOS ARTÍCULOS QUE ES AQUELLA INFORMACIÓN PROPIA DE LOS OBJETOS COMO EL NOMBRE O GÉNERO DEL ARTISTA O CANCIÓN.

10. ¿CUÁLES SON LAS TÉCNICAS MÁS USADAS?

LAS DOS TÉCNICAS MÁS USADAS SON EL FILTRADO COLABORATIVO Y EL FILTRADO BASADO EN CONTENIDO. EN EL FILTRADO COLABORATIVO SE BASA EN UNA MATRIZ USUARIO–ITEM Y EN EL HISTORIAL DEL USUARIO O USUARIOS SIMILARES, MIENTRAS QUE EL FILTRADO BASADO EN CONTENIDO SE BASA EN LAS SEMEJANZAS ENTRE LOS ARTÍCULOS, POR EJEMPLO, DOS TRAMAS DE DOS LIBROS MUY PARECIDAS .

11. ¿QUÉ MODELOS DE RECOMENDACIÓN HAY?

LOS MODELOS DE RECOMENDACIÓN QUE HAY SON DOS, BASADO EN MEMORIA Y BASADO EN MODELO. EL MODELO BASADO EN MEMORIA SE BASA EN LAS INTERACCIONES PASADAS SIN ASUMIR UN MODELO PREDEFINIDO, SUELE MIRAR LOS VECINOS MÁS CERCANOS, MIENTRAS QUE EL BASADO EN MODELO SUBYACENTE TIENE PREDEFINIDO POR DEBAJO UN MODELO QUE EXPLICA LA INTERACCIÓN ENTRE EL USUARIO Y EL ÍTEM.

12. ¿QUÉ SISTEMAS DE CALIFICACIÓN USUARIO/OBJETO HAY?

LOS DOS SISTEMAS DE CALIFICACIÓN OBJETO USUARIO SON EL EXPLÍCITO Y EL IMPLÍCITO. EL EXPLÍCITO ES AQUEL EN EL QUE EL USUARIO HACE UNA RESEÑA, DA UN FEEDBACK O ESTABLECE UNA PUNTUACIÓN A CIERTO SERVICIO, PRODUCTO O ARTÍCULO, MIENTRAS QUE EL SISTEMA DE CALIFICACIÓN IMPLÍCITO ES INFORMACIÓN QUE DERIVA DE LAS INTERACCIONES COMO POR EJEMPLO EL NÚMERO DE HORAS JUGADAS A UN VIDEOJUEGO O LOS CLICKS, LOS ELEMENTOS QUE METEMOS EN LA CESTA DE LA COMPRA O LOS QUE METEMOS Y NO LLEGAMOS A COMPRAR.

13. ¿CUÁLES SON LOS RETOS DE LOS SISTEMAS DE RECOMENDACIÓN?

LOS RETOS DE LOS SISTEMAS DE RECOMENDACIÓN SON QUE LAS CALIFICACIONES EXPLÍCITAS SON DIFÍCILES DE PROCESAR, ADEMÁS DEL PROBLEMA DE ARRANQUE EN FRÍO, ¿CÓMO EMPEZAMOS UN SISTEMA DE RECOMENDACIÓN EFICIENTE CON SOLO UNOS POCOS USUARIOS? Y POR ÚLTIMO, LA ESCALABILIDAD.

14. ¿CUÁLES SON LOS PASOS PARA EL PROCESAMIENTO DE DATOS CON MACHINE LEARNING?

EN PRIMER LUGAR SE CARGAN LOS DATOS Y SE LIMPIAN, DESPUÉS SE LIMPIAN LOS DATOS, SE APLICA INGENIERÍA DE CARACTERÍSTICAS (FEATURE ENGINEERING), SE ENTRENA EL MODELO Y POR ÚLTIMO SE EVALUA.

15. ¿CUÁLES SON LOS CONCEPTOS DE APACHE SPARK PARA EL PROCESAMIENTO CON MACHINE LEARNING?

LOS CONCEPTOS PRINCIPALES SON LA CARGA DE DATOS Y SU LIMPIEZA, EL TRANSFORMADOR EL ESTIMADOR Y EL EVALUADOR.

16. ¿QUÉ ES EL APRENDIZAJE SUPERVISADO? EXPLICA UN PROCESO MACHINE LEARNING CON APRENDIZAJE SUPERVISADO CON SPARK Y DA UN POSIBLE CASO DE USO.

EL APRENDIZAJE SUPERVISADO ES UNA TÉCNICA DE APRENDIZAJE AUTOMÁTICO DONDE EL ALGORITMO APRENDE DE UN CONJUNTO DE DATOS ETIQUETADOS CON EL OBJETIVO DE PREDECIR LAS ETIQUETAS DE UNOS NUEVOS DATOS SIN ETIQUETAR. LOS DATOS DE SALIDA PUEDEN SER DISCRETOS O CONTINUOS Y TENEMOS DOS TIPOS: CLASIFICACIÓN Y REGRESIÓN. EL PROCESO DE MACHINE LEARNING CON SPARK EMPIEZA CON LOS DATOS HISTÓRICOS ETIQUETADOS, QUE REPRESENTAN LA VARIABLE DEPENDIENTE. SE ENTRENA EL MODELO CON PARTE DE LOS DATOS ETIQUETADOS PARA QUE EL MODELO APRENDA PATRONES DE LOS DATOS. SE PASAN LOS DATOS DE

PRUEBA AL MODELO PARA VER QUÉ TAN PRECISO ES EN SUS PREDICCIONES. UN EJEMPLO PUEDE SER PREDECIR EL SUELDO DE UNA PERSONA (VARIABLE DEPENDIENTE) BASÁNDONOS EN SU EDAD (FEATURE) .

17. EXPLICA EL PROCESO DE ENTRENAMIENTO DE UN MODELO DE MACHINE LEARNING, EXPLICA CON QUÉ EMPIEZA EL PROCESO Y CON QUÉ FINALIZA. ¿POR QUÉ ES TAN IMPORTANTE EL PROCESO DE ENTRENAMIENTO?

EL PROCESO DE ENTRENAMIENTO DE UN MODELO GENERALMENTE SE DESARROLLA A TRAVÉS DE UN PROCESO ITERATIVO COMO UN ALGORITMO DE OPTIMIZACIÓN COMO EL DESCENSO DEL GRADIENTE. EL ALGORITMO DE ENTRENAMIENTO COMIENZA CON UN MODELO BÁSICO Y LO MEJORA GRADUALMENTE AJUSTANDO PARÁMETROS INTERNOS (COEFICIENTES) EN LAS ITERACIONES DEL ENTRENAMIENTO. EL RESULTADO DEL ENTRENAMIENTO ES UN MODELO ENTRENADO QUE PUEDE REALIZAR PREDICCIONES. EL PROCESO DE ENTRENAMIENTO ES TAN IMPORTANTE PORQUE EL MODELO APRENDE LOS PATRONES DE LOS DATOS PARA PODER GENERALIZAR A CUALQUIER CONJUNTO DE DATOS CON LAS MISMAS CARACTERÍSTICAS REDUCIENDO EL SESGO ENTRE LA PREDICCIÓN Y LA REALIDAD.

18. EXPLICA QUÉ ES UN MODELO O PROBLEMA DE REGRESIÓN (QUÉ OBJETIVO PREDICTIVO TIENE)

EL MODELO DE REGRESIÓN ES UN PROBLEMA DE MACHINE LEARNING EN EL QUE LA CARACTERÍSTICA A PREDECIR ES CONTINUA Y SE PREDICEN VALORES QUE EL MODELO NO HA VISTO DURANTE EL ENTRENAMIENTO.

19. EXPLICA QUÉ ES UN MODELO O PROBLEMA DE CLASIFICACIÓN (QUÉ OBJETIVO PREDICTIVO TIENE)

EL MODELO DE CLASIFICACIÓN ES UN PROBLEMA EN EL QUE SE ENTRENA UN ALGORITMO PARA PREDECIR UNA VARIABLE CATEGÓRICA. BÁSICAMENTE SE CLASIFICA EL ELEMENTO EN UNA DE LAS POSIBLES CATEGORÍAS DE LA VARIABLE DEPENDIENTE.

20. NOMBRA CINCO ALGORITMOS DIFERENTES DISPONIBLES EN SPARK MLIB Y QUÉ FUNCIONALIDAD APORTAN (CLASIFICACIÓN O REGRESIÓN)

REGRESIÓN LINEAR PARA REGRESIÓN, REGRESIÓN LOGÍSTICA PARA CLASIFICACIÓN, DECISIÓN ÁRBOL, RANDOM FOREST PARA REGRESIÓN Y CLASIFICACIÓN, Y NAIVE BAYES SVM PARA CLASIFICACIÓN.

21. ¿QUÉ DATOS DE ENTRADA RECIBE UN SISTEMA DE RECOMENDACIÓN? ¿QUÉ DATOS DE SALIDA CALCULA EL SISTEMA DE RECOMENDACIÓN?

PRIMERO SE RECIBEN LOS DATOS DE ENTRADA QUE CONTIENEN PREFERENCIAS IMPLÍCITAS O EXPLÍCITAS DE LOS USUARIOS PARA VARIOS ÍTEMS. EL ALGORITMO HACE RECOMENDACIONES DE QUÉ LE GUSTARÁ AL USUARIO ENCONTRANDO SIMILITUDES ENTRE USUARIOS O ARTÍCULOS. LOS DATOS DE SALIDA DE UN SISTEMA DE RECOMENDACIÓN SON LAS PROPIAS RECOMENDACIONES COMO LO QUE LE HA GUSTADO A USUARIOS SIMILARES O QUÉ OTROS PRODUCTOS SE PARECEN A LOS QUE EL USUARIO YA HA COMPRADO.

22. EXPLICA PASO A PASO EL FLUJO DE TRABAJO DE EL PROCESAMIENTO DE DATOS CON MACHINE LEARNING Y QUÉ ESTRUCTURAS INTERACTÚAN EN EL FLUJO EN SPARK.

1. SE RECOLECTAN LOS DATOS RELEVANTES PARA LA TAREA

2. SE LIMPIAN Y SE INSPECCIONAN LOS DATOS PARA COMPRENDERLOS MEJOR

3. SE REALIZA INGENIERÍA DE CARACTERÍSTICAS PARA PERMITIR QUE EL ALGORITMO APROVECHE LOS DATOS EN UNA FORMA ADECUADA.

4. USAR UNA PARTE DE ESTOS DATOS COMO CONJUNTO DE ENTRENAMIENTO PARA ENTRENAR UNO O MÁS ALGORITMOS PARA GENERAR ALGUNOS MODELOS CANDIDATOS.

5. EVALUAR Y COMPARAR MODELOS CON SUS CRITERIOS DE ÉXITO MEDIANTE MEDIR OBJETIVAMENTE LOS RESULTADOS EN UN SUBCONJUNTO DE LOS MISMOS DATOS QUE NO SE UTILIZARON PARA LA CAPACITACIÓN (TRAINING)

6. UTILIZAR EL MODELO PARA HACER PREDICCIONES DETECTAR ANOMALÍAS O DESHACER DESAFÍOS COMERCIALES MÁS GRANDES.

24. ¿QUÉ PUEDE PASAR SI CONSTRUIMOS UN MODELO CON LOS DATOS ENTEROS EN VEZ DE PARTICIONARLOS POR EJEMPLO EN 80-20?

PUEDE SER QUE EL MODELO MEMORICE LOS DATOS Y SE PRODUZCA OVERFITTING QUE QUIERE DECIR QUE PREDICE MUY BIEN LOS DATOS CON LOS QUE SE ENTRENA EL MODELO PERO ES INCAPAZ DE GENERALIZAR CON NUEVOS DATOS.

25. EXPLICA LA ESTRUCTURA: TRANSFORMADORES. PON UN EJEMPLO DE TRANSFORMADOR.

LOS TRANSFORMADORES EN SPARK ACEPTAN UN DATAFRAME COMO ENTRADA Y DEVUELVEN UN NUEVO DATAFRAME GENERALMENTE CON UNA NUEVA COLUMNA AGREGADA. NO APRENDEN DE SUS DATOS SINO QUE APLICAN TRANSFORMACIONES BASADAS EN REGLAS USANDO EL MÉTODO TRANSFORM. SE UTILIZA EN EL PREPROCESAMIENTO Y EN INGENIERÍA DE CARACTERÍSTICAS. POR EJEMPLO, EL VECTORASSEMBLER TRANSFORMER QUE ES UN TIPO DE TRANSFORMACIÓN DE APACHE SPARK EN EL QUE SE COMBINAN LOS VALORES DE DIFERENTES CARACTERÍSTICAS O COLUMNAS Y SE AGRUPAN EN UN VECTOR EN UNA NUEVA COLUMNA.

26. EXPLICA LA ESTRUCTURA: ESTIMADORES. PON UN EJEMPLO DE ESTIMADOR.

LOS ESTIMADORES APRENDEN PARÁMETROS DE NUESTROS DATOS, TOMAN UN DATAFRAME Y DEVUELVEN UN MODELO, COSA QUE REQUIERE DOS PASADAS SOBRE NUESTROS DATOS: LA PRIMERA PASADA PARA GENERAR VALORES DE INICIACIÓN MIENTRAS QUE LA SEGUNDA PASADA APLICA REALMENTE LA FUNCIÓN GENERADA. UN EJEMPLO ES LA REGRESIÓN LINEAR.

EN LA NOMENCLATURA DE SPARK A LOS ALGORITMOS QUE PERMITEN A LOS USUARIOS ENTRENAR UN MODELO A PARTIR DE DATOS TAMBIÉN SE LLAMAN ESTIMADORES.

27. EXPLICA LA ESTRUCTURA: EVALUADORES. PON UN EJEMPLO DE EVALUADOR

UN EVALUADOR NOS PERMITE VER COMO FUNCIONA UN MODELO DETERMINADO SEGÚN LOS CRITERIOS QUE ESPECIFICAMOS. SE SUELE USAR UN EVALUADOR PARA COMPROBAR LA CALIDAD DEL MODELO Y LUEGO SE DECIDE USAR ESE MODELO PARA HACER PREDICCIONES. EJEMPLOS DE MÉTRICAS PARA LOS EVALUADORES SON EL R^2 Y EL RMSE Y UN EJEMPLO DE EVALUADOR ES EL DE REGRESIÓN LINEAL (`REGRESIONEVALUATOR`).

28 EXPLICA LA ESTRUCTURA: PIPELINES.

PODEMOS ESPECIFICAR CADA UNA DE LAS TRANSFORMACIONES, ESTIMACIONES Y VALORACIONES UNA A UNA PERO A MENUDO ES MÁS FÁCIL ESPECIFICAR NUESTROS PASOS COMO ETAPAS DE UN PROCESO.

29. ¿QUÉ ES ALS?

ES UN TIPO DE ESTIMADOR QUE TOMA UN DATAFRAME Y DEVUELVE UN MODELO YA CONSTRUIDO QUE SE SUELE USAR PARA RECOMENDAR ÍTEMS CON SPARK.