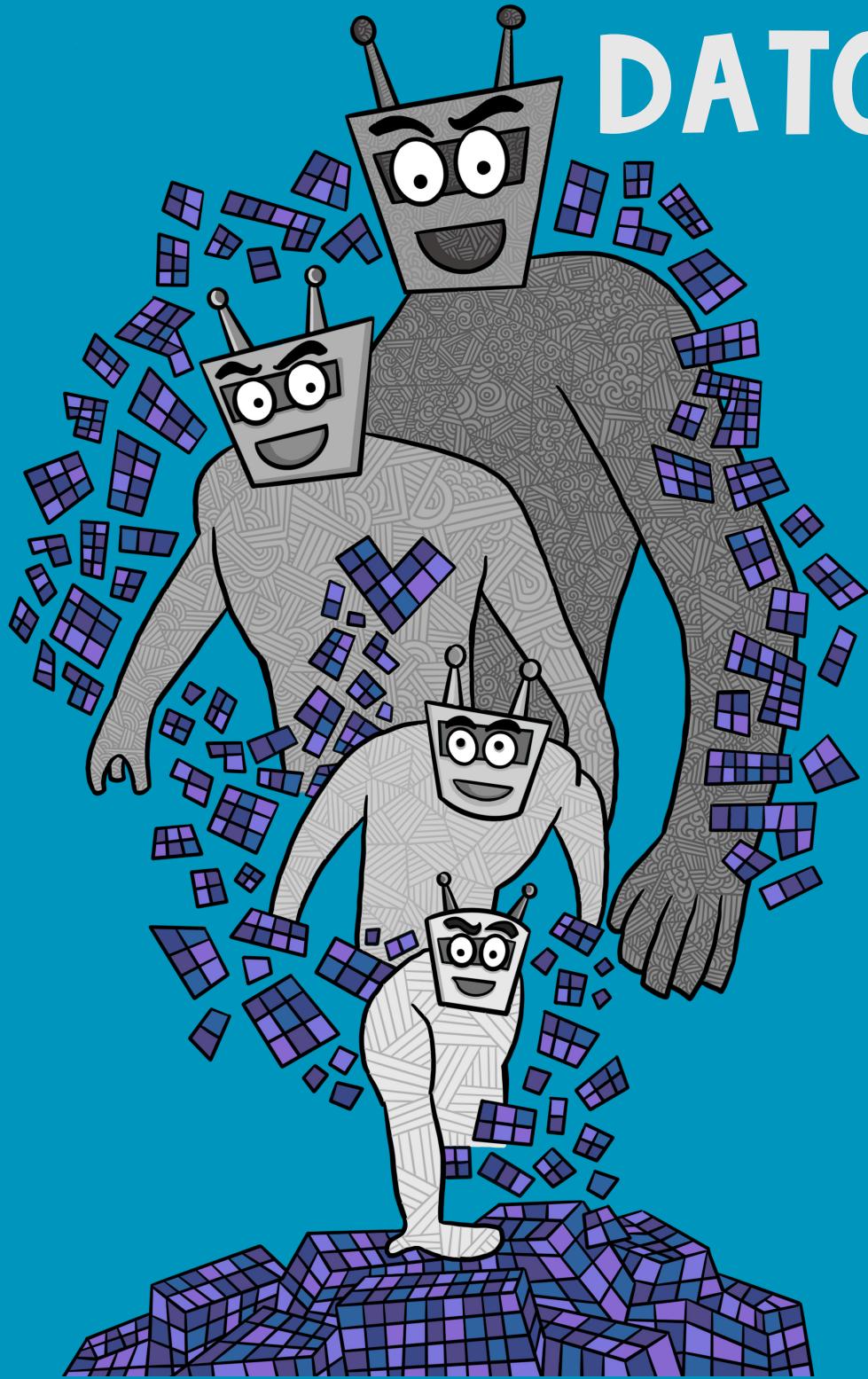


Somos IA #2:

APRENDER DE LOS DATOS



TERMINOS DE USO

Todos los paneles de este libro de historietas tienen licencia CC BY-NC-ND 4.0. Consulte la página de licencias para obtener detalles sobre cómo puede usar este material gráfico.

TL;DR: Se puede usar paneles/grupos de paneles en presentaciones/artículos, siempre y cuando

1. Se proporcione la cita adecuada
2. No se realicen modificaciones a los paneles individuales

Citar como:

Julia Stoyanovich y Falaah Arif Khan. “Aprender de los datos”. Somos IA, Vol. 2 (2022) https://dataresponsably.github.io/we-are-ai/comics/vol2_es.pdf

Contacto:

Dirija cualquier consulta sobre el uso de elementos de este cómic a themachinelearnist@gmail.com, con copia a stoyanovich@nyu.edu

CUANDO HABLAMOS DE INTELIGENCIA ARTIFICIAL (IA), GENERALMENTE HACEMOS UNA DISTINCIÓN ENTRE LOS SISTEMAS DE “IA CLÁSICA”:

AQUELLOS QUE FUNCIONAN DE ACUERDO CON REGLAS PREDETERMINADAS, COMO ROOMBA, LA ASPIRADORA INTELIGENTE,



Y AQUELLOS QUE “APRENDEN” LAS REGLAS A PARTIR DE LOS DATOS:

COMO LA IA PARA JUGAR AL AJEDREZ

O UN COCHE AUTÓNOMO.

PERO ESTA DISTINCIÓN NO ES ESTRICTA.

DE HECHO, LA MAYORÍA DE LOS SISTEMAS DE IA QUE EXISTEN EN EL MUNDO HOY EN DÍA COMBINAN REGLAS HECHAS A MANO CON ALGUNA FORMA DE APRENDIZAJE.

VEAMOS EJEMPLOS DE TALES REGLAS Y VEAMOS CÓMO LAS MÁQUINAS LAS APRENENDEN DE LOS DATOS.

SUPONGAMOS QUE TENEMOS QUE DISEÑAR UN SISTEMA DE ILUMINACIÓN INTELIGENTE



ESTE ALGORITMO ES MUY SIMPLE, CUENTA CON UN SOLO UN PASO:

TOMA LAS CONDICIONES EXTERIORES (SI ESTÁ OSCURO AFUERA) COMO ENTRADA,

Y, UTILIZANDO LA REGLA 1, PREDICE LA SALIDA ADECUADA: ENCENDER LA LUZ O NO.

ESTA REGLA ES SIMPLE, PERO ¿ES BUENA?

¿CAPTURA ADECUADAMENTE LA FORMA EN QUE OPERAMOS MANUALMENTE LAS LUCES EN CASA?

¡HAGAMOS UN EXPERIMENTO PARA AVERIGUARLO!

RECOGEREMOS OBSERVACIONES SOBRE SI ESTÁ OSCURO AFUERA Y SI LAS LUCES ESTÁN ENCENDIDAS EN LA CASA.

A

PARA CADA OBSERVACIÓN, COMPROBAREMOS SI LA PREDICCIÓN REALIZADA POR NUESTRO ALGORITMO COINCIDE CON LO QUE OBSERVAMOS.

PARA LAS OBSERVACIONES A Y B, ESTÁ OSCURO AFUERA, PERO ESTAMOS VIENDO LA TELEVISIÓN Y CENANDO,



ASÍ QUE LAS LUCES ESTÁN ENCENDIDAS.

¡Y ESTO ES, DE HECHO, LO QUE HA PREDICHO LA REGLA 1!

ESTO SIGNIFICA QUE TANTO A COMO B RESPALDAN LA HIPÓTESIS CODIFICADA POR LA REGLA 1.

B

LA OBSERVACIÓN C TAMBIÉN APOYA LA HIPÓTESIS:

ESTÁ DESPEJADO Y SOLEADO, ASÍ QUE LAS LUCES ESTÁN APAGADAS.

C

VEAMOS AHORA LA OBSERVACIÓN D:

ESTÁ OSCURO AFUERA Y ES HORA DE DORMIR; POR TANTO, LAS LUCES ESTÁN APAGADAS.

D

PERO, SEGÚN LA REGLA 1, EL ALGORITMO HA PREDICHO QUE LAS LUCES ESTARÍAN "ENCENDIDAS". ¡LA REGLA 1 FALLA AQUÍ!

VEAMOS LO QUE HEMOS APRENDIDO SOBRE LA REGLA 1 DE NUESTRO EXPERIMENTO:

NUESTRA REGLA PUDO PREDICIR CORRECTAMENTE LA SALIDA 3 DE 4 VECES.



HAGO ALGUNOS ARREGLOS EN LA TERMINOLOGÍA AQUÍ:

ES HABITUAL REFERIRSE A LA ENTRADA DE LA REGLA (SI ESTÁ OSCURO, EN NUESTRO EJEMPLO) COMO LA CARACTERÍSTICA DE ENTRADA, O SIMPLEMENTE **CARACTERÍSTICA**.

EL RESULTADO, EL ESTADO DE NUESTRAS LUces (ENCENDIDO/APAGADO), GENERALMENTE SE DENOMINA RESULTADO O ETIQUETA.



LA REGLA CON LA QUE ESTAMOS TRABAJANDO SE LLAMA "**CLASIFICADOR**": ASIGNA UNA ETIQUETA DE CLASE A UNA OBSERVACIÓN.

EL CLASIFICADOR QUE DISEÑAMOS USANDO LA REGLA 1 PUEDE HACER UNA DE DOS OPCIONES: ENCENDER O APAGAR LAS LUces.

TALES CLASIFICADORES SE LLAMAN "**BINARIOS**".

VOLVAMOS AHORA A LA OBSERVACIÓN D, QUE LA REGLA 1 CLASIFICÓ INCORRECTAMENTE:

¿CUÁL ES EL COSTO DE ESTE ERROR?

PUES, QUE LA LUZ INTELIGENTE NOS DESPIERTE EN MEDIO DE LA NOCHE.

¿QUÉ PODEMOS HACER AL RESPECTO?

SI COMPARAMOS LAS OBSERVACIONES A, B Y D,

VEMOS QUE TIENEN EL MISMO VALOR DE LA CARACTERÍSTICA DE ENTRADA (ESTÁ OSCURO FUERA), PERO SALIDAS DIFERENTES.

ESTO NOS HACE PENSAR QUE NECESITAMOS ALGUNAS CARACTERÍSTICAS ADICIONALES PARA DISTINGUIR ENTRE ESTAS SITUACIONES.

AHORA BIEN, UNA PREGUNTA QUE, COMO DISEÑADORES DE ESTE SISTEMA, DEBEMOS HACERNOS ES:

¿QUÉ OTRAS ENTRADAS PUEDEN SER ÚTILES?

SEGÚN SU EXPERIENCIA, ¿QUÉ CARACTERÍSTICAS CREE QUE SERÁN ÚTILES PARA PREDECIR EL RESULTADO?

¿LA TEMPERATURA EXTERIOR? ¡NO!

¿EL PRECIO DEL TÉ EN CHINA? ¡NO!

QUIZÁS, ¿ES HORA DE ACOSTARSE?

PONGÁMONOS A TRABAJAR PARA REFINAR NUESTRA REGLA.

AQUÍ ESTÁ NUESTRA REGLA REFINADA:

LA LLAMAREMOS REGLA 2.



"ENCENDER LAS LUCES SI ESTÁ OSCURO AFUERA,



DE LO CONTRARIO, MANTENER LAS LUCES APAGADAS."

Y AÚN NO ES HORA DE ACOSTARSE.

VOLVAMOS A EJECUTAR NUESTRO EXPERIMENTO, RECOPELEMOS ALGUNAS OBSERVACIONES MÁS Y EVALUEMOS NUESTRA NUEVA REGLA.

E



¡YA ESTÁ! ¡PARECE QUE FUNCIONA PERFECTAMENTE!

F

NO NECESITAMOS LAS LUCES ENCENDIDAS A MENOS QUE ESTÉ OSCURO AFUERA Y ESTEMOS DESPIERTOS HACIENDO ALGO!

G



H



Y ESTA VEZ, INCLUSO, EL GENIO APAGA LAS LUCES SI NOS QUEDAMOS DORMIDOS EN EL SOFÁ (DESPUÉS DE LA HORA DE ACOSTARNOS), ¡QUÉ CARIÑOSO!

LA REGLA 2 HA DEMOSTRADO UNA PRECISIÓN PERFECTA:

NUESTRO ALGORITMO ES CAPAZ DE PREDICIR EL RESULTADO CORRECTAMENTE EL 100 % DE LAS VECES.



HEMOS UTILIZADO NUESTRA INTUICIÓN Y EXPERIENCIA PARA DISEÑAR UNA REGLA ADECUADA, DATOS RECOPILADOS PARA VALIDAR SU RENDIMIENTO,

E HICIMOS LOS AJUSTES NECESARIOS HASTA QUE ALCANZAMOS UN DESEMPEÑO CON EL QUE ESTAMOS CONTENTOS.

DE HECHO, ESTAMOS MUY CONTENTOS CON NUESTRA LUZ INTELIGENTE,

LO CONVENIENTE QUE ES Y CÓMO NOS AYUDA A CONSERVAR LA ENERGÍA,

QUE QUEREMOS OFRECÉRSOLA A OTROS PARA QUE LA USEN.



VEAMOS SI FUNCIONA PARA QUE NUESTRA AMIGA ANN LO USE EN SU OFICINA.

PERO ANTES DE SEGUIR ADELANTE Y VOLVER A CABLEAR TODO EL EDIFICIO DE OFICINAS DE ANN, RECOPILEMOS ALGUNOS PATRONES DE USO PARA VER SI NUESTRA REGLA SERÍA EFECTIVA.

YA CONOCEMOS EL EJERCICIO: ¡HAGAMOS UN EXPERIMENTO!

RECORDEMOS LA REGLA 2:

"ENCENDER LAS LUCES SI ESTÁ OSCURO AFUERA Y TODAVÍA NO ES LA HORA DE ACOSTARSE. DE LO CONTRARIO, MANTENER LAS LUces APAGADAS".

I



EN LA OFICINA, LAS LUces ESTÁN ENCENDIDAS DURANTE EL DÍA CUANDO LAS PERSONAS TRABAJAN.

PERO LA REGLA 2 LAS APAGARÍA INCORRECTAMENTE, YA QUE HAY LUZ AFUERA.

J



LA MAYORÍA DE LAS NOCHES, CUANDO YA TODOS SE HAN IDO A CASA, LAS LUces SE APAGAN.

K



PERO LA REGLA 2 LAS ENCENDERÍA INCORRECTAMENTE, YA QUE ESTÁ OSCURO!

L

LOS FINES DE SEMANA, LA OFICINA ESTÁ VACÍA Y LAS LUces ESTÁN APAGADAS DURANTE TODO EL DÍA.

EN ESTE CASO, LA REGLA 2 FUNCIONA CORRECTAMENTE: MANTIENE LAS LUces APAGADAS, YA QUE ES DE DÍA AFUERA.

CADA CIERTO TIEMPO, ALGUIEN SE QUEDA HASTA TARDE EN LA OFICINA, TERMINANDO LAS ENTREGAS PARA UNA FECHA LÍMITE IMPORTANTE.

NECESITAMOS LAS LUces ENCENDIDAS EN ESTA SITUACIÓN, PERO LA REGLA 2 FALLA DE NUEVO: APAGARÍA LAS LUces, YA QUE ES TARDE (Y YA HA PASADO LA HORA DE ACOSTARSE).

¡OH, NO! ESO SIGNIFICA QUE SOLO OBTUVIMOS 1 PREDICCIÓN CORRECTA DE 4.

SI SUPONEMOS QUE LAS LUces ESTÁN ENCENDIDAS APROXIMADAMENTE LA MITAD DEL TIEMPO, BASTA CON LANZAR UNA MONEDA AL AIRE PARA DECIDIR SI ENCENDER LAS LUces O NO.

¿POR QUÉ LA PRECISIÓN DE LA REGLA 2 ES TAN BAJA?

PORQUE EL CONJUNTO DE DATOS EN EL QUE PROBAMOS EL CLASIFICADOR REPRESENTA UNA SITUACIÓN DIFERENTE A LA DEL CONJUNTO DE DATOS EN EL QUE LO ENTRENAMOS.

ENTRENAMOS AL CLASIFICADOR, DECIDIMOS LAS REGLAS APROPIADAS, EN FUNCIÓN DE NUESTROS PATRONES DE CONSUMO EN EL HOGAR,

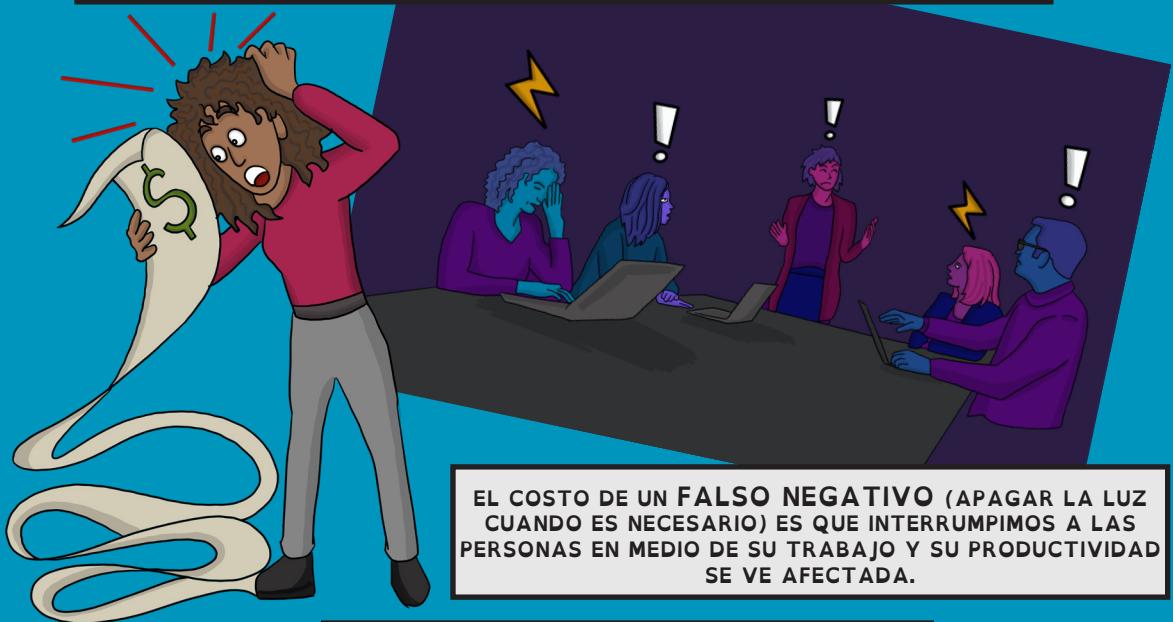
PERO ANN LO "PROBÓ" EN SU OFICINA.

YA SEA A LA HORA DE ACOSTARSE O SI ESTÁ OSCURO AFUERA, IYA NO IMPORTA MUCHO!

LO QUE IMPORTA ES SI LA OFICINA ESTÁ ABIERTA.

¿CUÁL ES EL COSTO DE UN ERROR AQUÍ?

EL COSTO DE UN FALSO POSITIVO (ENCENDER LA LUZ INNECESARIAMENTE CUANDO DEBERÍA ESTAR APAGADA) ES QUE DESPERDICIAMOS ENERGÍA Y UNA FACTURA DE ELECTRICIDAD CONSIDERABLE PARA LA COMPAÑÍA.



EL COSTO DE UN FALSO NEGATIVO (APAGAR LA LUZ CUANDO ES NECESARIO) ES QUE INTERRUMPIMOS A LAS PERSONAS EN MEDIO DE SU TRABAJO Y SU PRODUCTIVIDAD SE VE AFECTADA.

¿QUÉ DEBEMOS HACER? ¿VOLVER AL TABLERO DE DIBUJO OTRA VEZ?

PENSAR EN QUÉ CARACTERÍSTICAS ADICIONALES USAR, RECOPLAR ESOS DATOS,

Y REALIZAR UN EXPERIMENTO PARA VERIFICAR SI NUESTRA REGLA REALMENTE FUNCIONA?

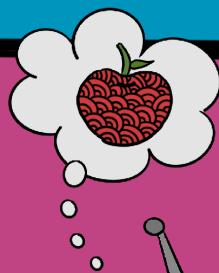
PERO ESTO ES TEDIOSO.

Y LO QUE ES MÁS IMPORTANTE, SERÁ MUY DIFÍCIL PARA NOSOTROS CONTINUAR AJUSTANDO ESTAS REGLAS QUE SON CADA VEZ MÁS COMPLEJAS,

SEGÚN LAS UBICACIONES DONDE SE UTILICE LA LUZ INTELIGENTE Y SUS DIFERENTES REQUISITOS.

LA RESPUESTA ES: USAR UN ALGORITMO DE APRENDIZAJE AUTOMÁTICO PARA IDENTIFICAR PATRONES ESTADÍSTICOS EN NUESTROS DATOS, PARA "APRENDER" LAS REGLAS DE FORMA AUTOMÁTICA.

PERO AL IGUAL QUE DISEÑAMOS EXPERIMENTOS PARA VERIFICAR SI LAS REGLAS QUE CREAMOS ERAN BUENAS,

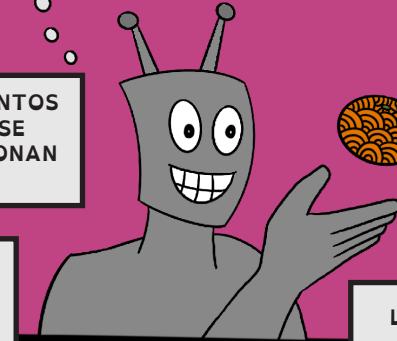


TAMBÍEN DEBEMOS DISEÑAR EXPERIMENTOS PARA PROBAR SI LAS REGLAS QUE SE APRENDEN AUTOMÁTICAMENTE FUNCIONAN CUANDO SE IMPLEMENTAN.

SI ESOS DATOS SON REPRESENTATIVOS DEL USO FUTURO DEL SISTEMA DE LUZ INTELIGENTE,

ENTONCES LA REGLA FUNCIONARÁ BIEN.

ESTO SE DEBE A QUE LA REGLA APRENDIDA DEPENDE COMPLETAMENTE DE LOS DATOS HISTÓRICOS:



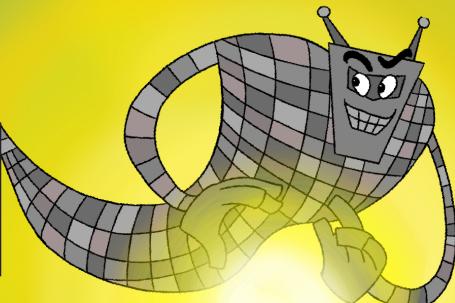
PERO SI EL CASO DE USO REAL ES DIFERENTE,

LA REGLA COMETERÁ MUCHOS ERRORES,

COMO ENCENDER LA LUZ EN EDIFICIOS DE OFICINAS VACÍOS,

O DESPERTAR A UN NIÑO PEQUEÑO QUE DUERME .

YA SEA QUE LAS REGLAS SEAN ESCRITAS POR HUMANOS O APRENDIDAS DE DATOS HISTÓRICOS POR MÁQUINAS, DEBEMOS ASEGURARNOS DE USAR EL MÉTODO CIENTÍFICO:



LA FORMULACIÓN Y PRUEBA DE UNA HIPÓTESIS FALSABLE.



COMO DIJO EL FILÓSOFO DE LA CIENCIA KARL POPPER:

“NINGUNA TEORÍA O IDEA DEBERÍA CONSIDERARSE CIENTÍFICA A MENOS QUE, EN PRINCIPIO, PUEDA DEMOSTRARSE SU FALSEDAD”.

EL MÉTODO CIENTÍFICO COMIENZA CON UNA OBSERVACIÓN.



OBSERVACION

LUEGO SIGUE LA FORMULACIÓN DE UNA HIPÓTESIS FALSABLE;

ES DECIR, QUE PUEDE DEMOSTRARSE COMO FALSA.



CONCLUSION

EL MÉTODO CIENTÍFICO



HIPOTESIS

Y, DE SER ASÍ, PERFECCIONAMOS NUESTRA HIPÓTESIS, DISEÑAMOS UN NUEVO EXPERIMENTO, ENJUAGAMOS Y REPETIMOS.



EXPERIMENTO

LUEGO, SE DISEÑA UN EXPERIMENTO PARA VERIFICAR SI LA HIPÓTESIS ES FALSIFICADA.

LA PREGUNTA CRUCIAL: "¿FUNCIONA EL CLASIFICADOR?" SE FORMALIZA COMO "¿LAS PREDICIONES DEL CLASIFICADOR SON MÁS PRECISAS QUE UNA CONJETURA ALEATORIA?"

ESTA ES LA VARA MÁS BAJA PARA LA PRECISIÓN:

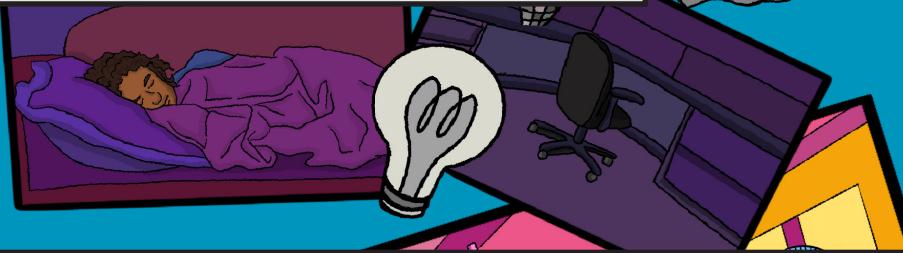


NO QUEREMOS TERMINAR CONSTRUYENDO UNA IA SOFISTICADA PARA LANZAR MONEDAS.

Y SI LA HIPÓTESIS ES FALSIFICADA,
NO LA USAMOS.

CUANDO TENEMOS
INFORMACIÓN VERDADERA
SOBRE EL PAPEL
(OBSERVACIONES DE SI LA
LUZ DEBERÍA O NO ESTAR
ENCENDIDA), PODEMOS
VERIFICAR.

DISEÑAMOS UN NUEVO PROCEDIMIENTO DE GENERACIÓN DE
DATOS Y VOLVEMOS A ENTRENAR EL MODELO PARA
APRENDER UNA NUEVA REGLA.



ES IMPORTANTE DESTACAR QUE, INCLUSO CON TODA ESTA SOFISTICACIÓN ADICIONAL,
ES POCO PROBABLE QUE ENCONTREMOS LA PRECISIÓN PERFECTA.



UNA DE LAS RAZONES POR LAS QUE ESTO OCURRE ES
QUE SIEMPRE HAY INCERTEZAS EN EL MUNDO.

OTRA ES QUE A VECES SE ROMPEN LAS REGLAS: LAS
PERSONAS PUEDEN VENIR A TRABAJAR DURANTE EL FIN DE
SEMANA O PERMANECER EN EL TRABAJO MÁS TIEMPO PARA
CUMPLIR CON UN PLAZO.



LO MEJOR QUE PODEMOS ESPERAR, EN
SITUACIONES DE LA VIDA REAL, ES QUE EL
CLASIFICADOR FUNCIONE LA MAYOR PARTE
DEL TIEMPO.

PERO DE VEZ EN CUANDO, ALGUIEN DEBERÁ LEVANTARSE Y
ENCENDER O APAGAR LA LUZ.