Instituto Politécnico Nacional

Licenciatura en Ciencia de Datos

Propuesta de normas para un protocolo que disminuya la discriminación racial algorítmica en la creación de algoritmos de Machine Learning a partir del análisis de algoritmos discriminatorios implementados en el pasado para su adaptación en México.

De Luna Ocampo Yanina

Sainz Takata Juan Pablo Minoru

Ramírez Méndez Kevin

Escuela Superior De Cómputo, Metodología de la investigación y divulgación científica.

Acosta Martínez Lilián

26 de noviembre de 2022

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Índice

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Introducción

La discriminación algorítmica se refiere a aquellos procesos a través de los cuales distintos tipos de discriminación que ocurren en el mundo real son reproducidos en entornos de datos o los que surgen exclusivamente en ellos, como cuando los sistemas de reconocimiento facial producen más errores al procesar rostros no caucásicos [1].

Y aunque pensemos que esto es tan fácil como dejar de considerar la raza como una variable para los cálculos algorítmicos es más complicado de lo que parece. Debido a esto, han empezado a pedir que los organismos manejen algoritmos que realicen un proceso justo y transparente, que expliquen la manera en cómo los sistemas automatizados toman decisiones, especialmente aquellas que afectan significativamente las vidas individuales [2].

Para terminar con este sesgo racial, dentro de los algoritmos, se ha encontrado que por medio de leyes existentes este se puede mitigar y que los sesgos son más difíciles de corregir si no tenemos diversidad en nuestra sociedad, mayormente en el campo laboral y en políticas públicas.

Para poder combatir esto, como se mencionó previamente, se ha planteado que se debe exigir transparencia no sólo en cuanto al desarrollo de algoritmos, sino también respecto a su uso y quiénes son los beneficiarios de dichos sistemas de toma de decisión, pues son una forma más en la que continúa la vigilancia hacia los usuarios. Esto aunado al hecho que existe una tendencia hacia el monopolio en estos contextos, liderados por Google, YouTube, Facebook. Estas compañías controlan casi en su totalidad la información a la que tenemos acceso y la forma en la que es presentada para nuestro consumo.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Si no se tiene diversidad en el grupo de investigación no se podrán reconocer los problemas a los que enfrenta la mayoría de la gente, cuando los problemas no nos afectan, no pensamos en ellos, no les damos la importancia o inclusive no somos conscientes de su acontecer, porque no interactuamos con personas que los experimenten.

Algunas medidas que se han puesto para desvanecer la discriminación algorítmica han sido:

Un grupo de investigadores formando el proyecto data nutrition, el cual busca crear herramientas y prácticas que alienten al desarrollo responsable de inteligencias artificiales. Este proyecto sugiere agregar metadatos a nuestro conjunto de datos que describen la calidad de estos.

Otro enfoque complementario es el uso de las mismas técnicas de machine learning para identificar y cuantificar sesgos en algoritmos y en datos. A este modus se le conoce como IA de auditoría, en la cual el auditor es un algoritmo que sistemáticamente prueba el modelo de machine learning original para identificar sesgos en ambos el modelo y los datos de entrenamiento.

Un ejemplo de este enfoque se puede observar al usar la técnica word embedding para cuantificar estereotipos históricos. Esta técnica mapea cada palabra a un punto en el espacio, tal que la distancia entre vectores captura similitudes semánticas entre palabras. Capturando de esta forma analogías, tales como hombre es a rey como mujer es la reina. En este caso se usó una IA de auditoría para consultar a la técnica de word embedding por otras analogías de género. Revelando de esta forma analogías tales como hombre es a doctor como mujer es a enfermera o hombre es a programador como mujer es a ama de casa. Una vez que el auditor revela estos estereotipos en el modelo principal es posible reducir este sesgo modificando el mapeo de los vectores de palabras.[3]

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Marco metodológico

Introducción

En esta sección se pretende esclarecer la forma en la que se abordará la investigación, describiendo el enfoque que guía la investigación, siendo este el óptimo para dar respuesta al problema de investigación planteado. Además, de especificar al objeto de estudio dentro de este trabajo, el instrumento de investigación con el que se recolecta la información y la forma en la que esta se analiza.

1. Tipo de investigación

Aquí se definirá la forma en la que la investigación toma su curso, corrientes, métodos y características de esta.

Esta investigación se basa en la corriente del empirismo lógico como directriz epistemológica dando de esta forma validez al objeto de estudio mediante observaciones, pues todas las teorías se pueden comprobar en términos de sensaciones y relaciones de estas sensaciones [4], dando lugar al uso del método inductivo permitiendo que de las observaciones por analizar se llegue a una conclusión para el tema de investigación. Asimismo, se sigue un paradigma metodológico cualitativo, pues, seguir este enfoque nos permite contextualizar el fenómeno y tener una inmersión en el tema de investigación más adecuada, además, de tener un proceso circular en el que la recolección de datos servirá para afinar las preguntas de investigación e inclusive descubrir nuevas interrogantes o procesos de interpretación.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Esta investigación no se fundamenta con base en la estadística y no se pretende seguir una secuencia lineal. Es una investigación cuyo propósito es buscar, analizar e interpretar información documental, obtenida de los diferentes artículos seleccionados para llevar a cabo este trabajo. Este método remite a una aproximación holística al fenómeno en estudio, pudiendo indagar en la perspectiva de los expertos investigados en la documentación y de los sujetos a entrevistar y cómo conciben los factores conceptuales importantes en este tema de investigación. De este modo su utilización permite un acercamiento a los discursos, las prácticas y las relaciones que se produzcan.

Ad hoc a los objetivos propuestos y el planteamiento del problema, esta investigación es de índole explicativa, consecuentemente, la investigación no sólo busca describir al fenómeno investigado, como también dar motivo y razón al mismo, observando correlaciones y situaciones de causa y efecto entre las variables. También, se toma en cuenta el alcance transversal donde se tiene como objeto de estudio documentos con cierta temporalidad, en un entorno con amplitud macro sociológica centrándose en la estructura social, los sistemas sociales y la población a gran escala.

2. Población y muestra

Para la presente investigación, la población objeto de estudio está conformada por algoritmos discriminatorios implementados. Para esta investigación el objeto de estudios se seleccionaron casos de discriminación algorítmica documentados alrededor del mundo, además se entrevistaron a expertos para obtener más datos.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

3. Técnicas e instrumentos para la recolección de datos

Tomamos como una técnica a todas las formas de obtener datos para su procesamiento y ser transformados en información y un instrumento es el medio que es empleado para la recolección de datos.

Técnica:

Se hará uso de la entrevista a expertos en el tema con la finalidad de esclarecer la ocurrencia del fenómeno estudiado, y tener opiniones válidas y variadas con respecto a las causas de ocurrencia del mismo objeto de estudio del tema de investigación, para con ello llegar a una respuesta. Haciendo énfasis en que el uso de la información recopilada en las entrevistas es meramente de uso divulgativo y analítico.

Instrumento:

Guion de Entrevista

Datos generales de la entrevista

Nombre del entrevistado:	Fecha:
Cargo:	Lugar:

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Datos Específicos de la Entrevista

¿A qué se dedica actualmente?
¿En qué proyectos y cómo ha implementado algoritmos de ML en su trabajo?
¿Cómo describiría en palabras propias la discriminación algorítmica?
¿Ha oído de algún caso de discriminación algorítmica?
¿Qué aspectos considera que lleva a implementar algoritmos discriminatorios?
En su opinión profesional, ¿cuáles deberían de ser las medidas a tomar para evitar este tipo de sesgos?

Validación

Se refiere al grado en que un instrumento mide la variable que pretende medir. Puede referirse a la validez de contenido, de criterio o de constructos (Hernández-Sampieri, 1991). Puede ser determinada mediante el juicio de expertos.

En este caso, el instrumento de investigación fue validado por José Martín Ávila Martell, ingeniero en machine learning y desarrollo de software, actualmente en Apple, con una trayectoria en empresas internacionales como Intel y Xerox a quien se le brinda un agradecimiento por la contribución sustancial al validar el diseño del instrumento de recopilación de datos, haciendo énfasis en un uso meramente divulgativo e informativo de los datos por recolectar.

4. Técnicas de análisis de resultados

El análisis de datos se realizará mediante la teoría fundamentada, la cual se comprende como una metodología que tiene por objetivo develar y desarrollar teorías emergentes a partir de los datos recolectados y analizados de manera sistémica e interrelacionada. Llevando a cabo un proceso de verificación de los textos a revisar, clasificando con respecto a criterios específicos como:

- El impacto que tuvo el algoritmo discriminador en ese caso
- Los parámetros que se siguieron en su implementación.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Marco teórico

Introducción

En esta sección se muestran los sustentos teóricos que fundamentan y rigen a la investigación desarrollada, explicando y dando contexto a los objetivos y a las preguntas de investigación que nos permiten tener una mejor inmersión en el tema de investigación.

En primera instancia se fundamentan las corrientes filosóficas y teorías empleadas en la metodología de la investigación y su relación con el tema de investigación.

Empirismo Lógico

El empirismo lógico es una corriente filosófica de la epistemología que encapsula lo válido y veraz del conocimiento en fenómenos observables, de modo que esta corriente se basa en la inducción para la generación de conocimiento. En esta investigación se pretende dar explicación al porqué de los algoritmos que discriminan, esto a partir del análisis de casos prácticos y opiniones de expertos, con el fin de poder generalizar las causas por las que existen dichos algoritmos.

Teoría Fundamentada

Es un método de investigación cualitativa que busca en los datos conceptualizaciones emergentes en patrones integrados y categorizados analizando, a través de pasos rigurosos, en un proceso de constante comparación. Este método está diseñado para generar conceptos y teorías que se fundamentan en los datos [5] Este método lleva al investigador a seguir un

paradigma comparativo donde se analice las similitudes y diferencias en los datos recopilados brindando así la factibilidad de categorizar las similitudes en los casos de discriminación algorítmica.

Discriminación

Según la Ley Federal para Prevenir y Eliminar la Discriminación (LFPED), la discriminación es toda distinción, exclusión, restricción o preferencia que, por acción u omisión, con intención o sin ella, no sea objetiva, racional ni proporcional y tenga por objeto o resultado obstaculizar, restringir, menoscabar o anular el reconocimiento, goce o ejercicio de los derechos humanos y libertades. Puede entenderse por discriminación el darle un trato desigual a una persona o grupo de personas en función de etiquetas que diferencian al individuo. Esta tiene lugar cuando no es posible usufructuar de nuestros derechos debido a una distinción injustificada. La discriminación disminuye o impide la libertad, agrede el sentido mismo de lo humano, disminuye la cohesión social, debilita las democracias, es tierra fértil para la violencia y obstaculiza el desarrollo [6].

En el fondo la discriminación está basada en prejuicios y patrones socioculturales aprendidos y repetidos, la conducta de las personas en nuestro entorno sociocultural juega un papel importante en el establecimiento de la identidad del individuo, es de aquí de donde nace el sentido de pertenencia. Esto genera una segmentación en la población que a su vez puede dar lugar a situaciones de odio, exclusión, desigualdad y deshumanización [6].

1. Discriminación en México

En México, el 20.2% de la población de 18 años y más declaró haber sido discriminada en el último año por alguna característica o condición personal, tono de piel, manera de hablar, peso o estatura, forma de vestir o arreglo personal, clase social, lugar donde vive, creencias religiosas, sexo, edad y orientación sexual. Al agrupar las categorías de tono de piel, peso o estatura y forma de vestir o arreglo personal, que forman parte de la apariencia de las personas, se observa que más de la mitad de la población en el rango de edad, ha percibido discriminación en el último año a estos rasgos.[7]

La presencia de estas prácticas discriminatorias en ámbitos institucionales específicos tiene como consecuencia efectos acumulativos, llevando a la discriminación estructural y desigualdad social, la cual se define como las acciones u omisiones de un Estado que a partir del no reconocimiento o del incumplimiento sistemático de derechos y libertades fundamentales de sectores de la población produce, reproduce o agrava desigualdades. Negar de manera injustificada un derecho, ya sea de forma directa o indirecta, es una situación que vulnera el derecho a la igualdad de las personas, produciendo resultados inequitativos para ciertos grupos sociales y reproduciendo la desigualdad social.[7]

Los derechos de los que la sociedad mexicana ha sido privada según su percepción son la atención médica, la atención a un servicio en una oficina de gobierno, la entrada o permanencia en algún negocio, en centros comerciales, al recibir apoyos de programas sociales, al obtener algún crédito de vivienda, préstamo o tarjeta. El 23.3% de la población de 18 años y más considera que en los últimos cinco años, se le negó injustificadamente algún derecho [7].

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Racismo

Uno de los casos más persistentes de discriminación social y en el que se enfoca esta investigación es en la discriminación racial, es por ello por lo que se presenta, describe y contextualiza este factor.

El Racismo es el odio, rechazo o exclusión de una persona por su raza, color de piel, origen étnico o su lengua, que le impide el goce de sus derechos. Es propiciado por un sentimiento irracional de superioridad de una persona sobre otra.

El Racismo se puede catalogar en tres diferentes escalas dependiendo del contexto en el que aparece, las principales categorías son:

- Racismo Institucional

Se da en acciones públicas que favorecen a una población y afectan o ignoran a otras.

- Racismo Sistémico

Se da cuando la sociedad en conjunto ve de forma inferior a un individuo o grupo por diferencias estereotipadas.

- Racismo Individual

Los prejuicios personales que nos hacen excluir o dar un trato desigual a un individuo dado su origen étnico, color de piel, etc.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

México está viviendo un proceso en contra de la discriminación racial, pero este aún es incipiente. La constitución ha reconocido el derecho a la no discriminación por origen étnico y ahora obliga legalmente a promover la igualdad de oportunidades de las personas indígenas en su artículo cuarto. Sin embargo, esto puede llegar a parecer irrelevante, considerando la desigualdad de trato real, pareciera normal los prejuicios y abusos, pues no podemos olvidar que la discriminación se basa en construcciones culturales y en un sistema de privilegios.

Como ejemplo de esto, el 40.3% de la población indígena declaró que se le discrimina debido a su condición de persona indígena. Los principales ámbitos donde se percibió esta discriminación fue en servicios médicos, la calle o transporte público y en la familia.[7]

Inteligencia Artificial (IA)

Es importante definir ciertos conceptos que guían la presente investigación, a continuación, se habla de la inteligencia artificial (IA) pues esta rama del conocimiento tiene una relevancia fundamental en el desarrollo de los algoritmos discriminatorios, objetos de estudio de esta investigación.

1. Algoritmos de machine learning

Hay diferentes definiciones acerca de qué son los algoritmos. Para contextualizar un poco, colocaremos algunas definiciones que encontramos.

La definición dada por la autora Pilar Rivas Vallejo sobre un algoritmo es una “secuencia finita de reglas formales que hacen posible obtener un resultado a partir de la entrada de la

información". Opina también que "un algoritmo es un conjunto de reglas que, aplicadas sistemáticamente a unos datos de entrada apropiados, resuelven un problema en un número finito de pasos elementales." [8]

La definición dada por la Real Academia Española nos dice que es un "conjunto ordenado y finito de operaciones que permite hallar la solución de un problema". [9]

Tomando estas dos definiciones, nosotros formulamos como algoritmos la siguiente, un algoritmo es una secuencia de pasos que resuelve un problema planteado.

Debemos tener en cuenta otro concepto para esta investigación y es, qué es el aprendizaje automático, de igual forma encontramos diferentes definiciones para este término.

La primera definición obtenida es dicha por la autora Pilar Rivas Vallejo, nos dice que es la "rama de la Inteligencia Artificial que tiene como objetivo desarrollar técnicas que permitan a las computadoras aprender." [8]

Como segunda definición, tomaremos la que ha planteado BBVA, que es "la rama de la Inteligencia Artificial que permite que las máquinas aprendan sin ser expresamente programadas para ello", añadiendo que, "es un maestro del reconocimiento de patrones, y es capaz de convertir una muestra de datos en un programa informático capaz de extraer inferencias de nuevos conjuntos de datos para los que no ha sido entrenado previamente". [10]

Tomando estas dos definiciones, nosotros formulamos como aprendizaje automático la siguiente, el aprendizaje automático lo entendemos como la parte de la Inteligencia Artificial que permite a un sistema aprender de los datos para así poder llegar a los objetivos.

2. Cómo se utilizan

Aquí debemos tener en cuenta algo importante, el algoritmo podemos diseñarlo para efectuar una selección, un ranking o hallar la mejor combinación de variables en función del resultado que se busca o se persigue. Teniendo presente que lo que suele definir al código fuente de las empresas es la propuesta en funcionamiento de los datos que los alimentan, estos son los que marcan el resultado y con ello el sesgo que se plantea. Estos algoritmos son empleados para tomar decisiones a partir de fuentes masivas de datos que se reciben, por lo tanto, podemos decir que del análisis que se obtiene, este se basa de la experiencia del anterior, por lo que se condiciona. Su capacidad de razonamiento está fundamentada en la réplica histórica, aun cuando se utilicen con un propósito predictivo.

De la misma manera, debemos tener en cuenta la cantidad de datos que se analiza ya que puede llegar a sesgar los datos, explicando este punto, tenemos que pensar que, si tenemos la mentalidad de que entre más datos metamos, el algoritmo será mejor, esto puede no ser cierto y la mayoría de las veces es así, por no decir que todas las veces, ya que podemos llegar a sobre entrenar el algoritmo por la cantidad de datos. La curva de nuestro algoritmo en vez de aprender de los datos ya obtenidos, lo que hará es que se aprenderá cada posibilidad como si fueran las únicas que podemos encontrar, y al momento de que le llegue algo ligeramente diferente, no sabrá cómo utilizar esa posibilidad, lo que hará que el algoritmo no sea funcional para los nuevos datos, solo servirá para los datos históricos.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Aplicaciones de los algoritmos

Los algoritmos están teniendo cada vez un papel más significativo en asuntos tan esenciales como los servicios públicos: desde la dotación de bonos sociales, la identificación de niveles de riesgo en situaciones de violencia de género, o la automatización de los sistemas de recursos humanos y contratación en empresas, hasta la decisión de cuántas patrullas policiales deben ir a cierto barrio en una ciudad porque se ha detectado que es una zona de mayor peligrosidad.

Esto permite optimizar la toma de decisiones, detectar patrones en situaciones críticas de forma más rápida o procesar cantidades masivas de información para dotar de una solución rápida. Sin embargo, por otro lado, la toma de decisiones basada en algoritmos, tal y como se está desplegando ahora, agrega más alcance, escala y velocidad a las consecuencias negativas de los sistemas tradicionales de distribución y reparto de servicios a la gente en situación de pobreza, a ello se une el hecho de que los modelos de lenguaje y de procesamiento de datos de los algoritmos incorporan sesgos de género y de raza [11].

1. Contextualizando el incremento del uso de algoritmos

El machine learning es una rama de la inteligencia artificial que permite que las máquinas aprendan de los datos en lugar de aprender de la programación explícita, por consiguiente, los algoritmos de machine learning tienen como característica el procesar grandes cantidades de datos históricos e identificar patrones de datos. Esto permite generar resultados de mayor precisión basados en los conjuntos de datos de entrada.

2. Usos comunes

La razón por la que el uso de modelos de machine learning tienen relevancia en cualquier industria es debido a que permite que la toma de decisiones empresariales sea basada en datos. Algunos ejemplos donde podemos observar las aplicaciones del machine learning son.

a. Medicina

La proliferación de dispositivos del área médica ha generado un importante volumen de datos sanitarios, con los cuales se ha apoyado al diagnóstico y tratamiento en tiempo real. Soluciones que detectan tumores cancerosos y diagnostican enfermedades oculares son algunos ejemplos de cómo ha impactado el machine learning en el área médica.

b. Marketing

Las empresas recurren al machine learning para comprender mejor a sus clientes y poder ofrecer productos personalizados o que sean más afines con ciertos sectores de la población.

c. Finanzas

El machine learning ha ayudado al sector bursátil y financiero, mediante el análisis de los movimientos del mercado, analizando riesgos e inclusive identificando clientes que puedan significar un préstamo de alto riesgo y mitigar los indicios de fraude.

Discriminación algorítmica

1. Casos donde se ha dado la discriminación algorítmica

Se ha demostrado que, con cambiar una palabra para el algoritmo, este presenta actitudes raciales, se hizo un estudio para demostrar este comportamiento, el resultado fue que al poner "el hombre blanco trabaja como…" la IA lo completaba como "un oficial de policía, un juez, un presidente", en cambio, si se escribía "el hombre negro trabaja como…" la IA lo completaba como "un proxeneta durante 15 días". [12]

Otra investigación realizada en Estados Unidos entrenó una IA con textos de internet, reveló que los nombres asociados con ascendencia europea, como Adam y Katie, tenían más probabilidades de ser vinculados con palabras agradables que los nombres asociados con ser afroamericanos como Alonzo o Latisha. [13]

Relacionado a esto, empresas grandes como lo es Google, han tenido que disculparse por situaciones en donde se han visto involucrados, ya que su algoritmo etiquetó erróneamente a las personas negras como "gorilas". [14]

Dentro de la medicina también han sucedido sesgos de importancia, en una investigación realizada y publicada por la revista Science mostró que un algoritmo que guiaba la atención médica de más de 200 millones de estadounidenses cada año, estaba sesgado en contra de las personas negras. Se estimó que este sesgo redujo el número de pacientes negros identificados para recibir programas de atención a más de la mitad. [15]

Se le preguntó a un periodista que trabaja en la antidiscriminación y en Inteligencia Artificial, qué lo llevó a crear AlgoRace, que es un equipo que busca de racionalizar la IA para poder

disminuir el impacto que tiene el racismo en la Inteligencia Artificial actualmente, a lo que contestó: “estamos inundados de algoritmos, están en todos los ámbitos de la vida. Y los hay positivos, pero la mayoría también reproducen sesgos racistas y de género.”, él afirma que el contexto del por qué se crean estos sesgos es porque el campo en el que se desarrollan analiza, diseñan y discuten los sistemas de inteligencia artificial, compuestos por algoritmos, están en un mundo masculinizado y blanco. Reafirma que, la funcionalidad de los algoritmos se ve afectada por los sesgos que tienen las personas que lo crean, lo prueban y los diseñan. [16]

Como se ha recalcado a lo largo de esta investigación, las propias series de datos que maneja el algoritmo son las que inducen a un resultado sesgado. Para esclarecer este punto, el investigador Dans menciona un ejemplo, que planteamos a continuación:

“Históricamente, en Estados Unidos la población negra ha tenido más problemas para devolver sus préstamos, lo cual es comprensible teniendo en cuenta la discriminación histórica que han sufrido, que los ha llevado a tener peor poder adquisitivo, etcétera”. Pues bien, aunque el banco en cuestión no guarde registro del origen étnico de sus clientes, “es posible que la estadística avanzada que usa el algoritmo llegue a la conclusión de que es mejor negar los préstamos a personas con unos nombres determinados, o que viven en unos barrios determinados o con cualquier característica que pueda estar correlacionada con la variable racial”. “En este caso, se estaría permitiendo que una serie histórica de datos hiciese que el algoritmo heredará un comportamiento determinado que consolida estereotipos raciales”. [17]

2. La dificultad de analizar los algoritmos para hacer hincapié en la causa del sesgo

Otro concepto importante dentro de nuestra investigación es el llamado “black box” o en también conocido como “epistemic opacity”, es el problema real cuando tienes que tomar decisiones en cuanto a la discriminación.

Al llegar a este punto, el algoritmo entrenado se convierte en una caja negra, donde resulta difícil conectar los resultados ofrecidos con los inputs recibidos, porque el propio algoritmo ha auto aprendido. Es decir, se conocen los datos que funcionan como input o estímulo de entrada, y se conoce el resultado o respuesta, pero el proceso intermedio permanece oculto, por lo que el humano no percibe cómo es que el algoritmo llegó a esa conclusión. Ya que, el acceso al código fuente no permite conocer realmente el origen de la decisión si el sesgo no está en su diseño sino en su alimentación por datos. Pero los resultados pueden advertirse como anómalos a fin de corregir el mecanismo, esto no salva a los ya afectados por su resultado sesgado, pero sí a las futuras entradas que existan. Aunque difícil no es imposible, diferentes técnicas permiten acceder a la arquitectura de decisión o, lo que es más importante, influir sobre ella para evitar ciertos sesgos en sus resultados, sin embargo, dentro de esto, si lo analizamos, estamos sesgando al sesgo. [18]

Como no podemos analizar mucho como el algoritmo funciona por lo antes mencionado, otra cosa que sí podemos hacer es analizar la información que le damos al algoritmo.

3. Falta o carencia de información

Se han buscado explicaciones del porque surge la discriminación algorítmica, Ignacio N. Cofone propone que este es un problema de información y “un algoritmo solo puede ser tan

bueno como la información que le damos" y que esta discriminación mayormente viene de "los datasets que involucran grupos que durante la historia han vivido desventajas sociales estos pueden sufrir de calidad de datos por varias razones, una de estas siendo la falta de representación"[19], por eso propone la adaptación de las leyes de privacidad ya que estas no funcionan para detener la discriminación algorítmica porque inconscientemente, aunque no le demos datos de raza, estos discriminan.

Otra postura sobre los datos, contraria a la creación de leyes de privacidad, es que la discriminación se crea por la ausencia de datos llamada discriminación estadística esta se ocasiona al no tener suficiente información para predecir un resultado, entonces al proponer leyes de privacidad tendríamos menos datos con los que trabajar y por ende la discriminación algorítmica aumentaría. Lior Strahilevitz argumenta que "al incrementar la disponibilidad de los datos sobre individuos, podemos reducir la dependencia que tienen los algoritmos que hacen decisiones acerca de grupos sociales "[19] y, por ende, "usualmente existe un conflicto esencial entre las protecciones de privacidad de la información y los principios antidiscriminatorios, tal que al reducir las protecciones de privacidad se reducirá el predominio de la discriminación estadística desagradable" [19].

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Entrevistas

En fin, de ampliar el análisis a los casos particulares de discriminación algorítmica, se recopila la opinión de expertos en el tema, quienes han ofrecido su punto de vista, con respecto a las implicaciones que tienen estos algoritmos y formas de disminuir los sesgos dentro de su implementación.

ENTREVISTA 1

Nombre del entrevistado: Zagal Flores Roberto Eswart	**Fecha:** 24 de noviembre del 2022
Cargo: Profesor Investigador	**Lugar:** Escuela Superior de Cómputo

¿A qué se dedica actualmente? Profesor investigador en la línea de datos de Ciencia de Datos para aplicaciones urbanas.
¿En qué proyectos y cómo ha implementado algoritmos de ML en su trabajo?

He trabajado en 2 proyectos de investigación en el CONACYT.

En el primero era para trabajar con patrones delictivos y ahí lo que hacíamos exploramos redes sociales: Facebook y hacíamos una integración de varios grupos que describen crímenes, en Ecatepec sobre todo, y ahí lo que hacíamos, limpiamos los datos sobre todo, y de ahí los íbamos procesando con una técnica que se llama "topic modeling" que es de las más tradicionales del lenguaje natural, lo que hace este tipo de técnicas es un algoritmo de clustering en texto, entonces lo que hace, va identificando grupos de publicaciones similares y ya después aplica otras técnicas para que pueda emerger los tópicos más comunes, o sea en este caso por ejemplo, puede ser un tópico compuesto por un bigrama o un eneagrama, que son cuando le pides al algoritmo que vaya buscando más tópicos con más de una palabra, entonces cuando emergen los tópicos vas a tener a los patrones para, por ejemplo, saber qué tipo de delitos es el que más se denuncia. O sea, caracterizan el delito y ahí es muy interesante por ejemplo es que conforme vamos bajando los datos podemos ir etiquetando y ahí lo que hacíamos nosotros era decir: "ah bueno, este es un robo de vehículo, este es un asalto, etc.", entonces lo que se podía hacer ahí, se podía automatizar el proceso, por ejemplo, están los datos de Facebook, lo metes a un clasificador y el clasificador puede decirte qué tipo de denuncia, cuando tienes el conjunto de denuncias, lo metes a un algoritmo de estos y te saca cuáles son los términos para este, eso fue lo que hicimos en el 2017 – 2019.

Y ya después, estuvimos viendo otro proyecto de salud, ese es más del tipo minería, big data, entonces era un ETL donde reconstruimos la información del sistema de salud de CDMX, y ya que teníamos toda la base de datos, se aplicaban unos estándares médicos, para por ejemplo, analizar un tipo de enfermedad, por ejemplo: diabetes y ya con eso se podía hacer una regresión lineal para decir por ejemplo, cuál era la probabilidad de que una persona pudiera fallecer o caer en hospital, dependiendo de

las combinaciones de afectaciones como sobre peso, edad, si es hipertensa, riesgo, etc., a partir de esos datos.

¿Cómo describiría en palabras propias la discriminación algorítmica?

Depende de los datos, o sea si el tipo de datos son objetos difícilmente vas a encontrar un aspecto de discriminación, pero si los datos son personas, dependiendo de la tendencia de cómo están distribuidas las características de estas personas, el algoritmo va a aprender y por ejemplo obviamente va a tomar una decisión con base en la tendencia de los datos, te pongo un caso, o sea imagínate que pudiéramos que automatizar el proceso de becas, entonces tú metes los datos del alumno, a lo mejor el algoritmo posiblemente pueda discriminar a una persona que posiblemente viva en Gustavo A. Madero, quizás en una colonia donde no hay índice de marginalidad medio pero a lo mejor no consideras otras variables, a lo mejor por la pandemia no tiene familiares y no le das beca, entonces tiene donde vivir pero a lo mejor necesita de lo básico para moverse, entonces eso depende del sesgo de los datos y depende también de la aplicación de lo que se está haciendo, se va a dar más dependiendo de los datos, las personas y dependiendo de la distribución de esas características.

¿Ha oído de algún caso de discriminación algorítmica?

Al menos en México todavía no, pero si puede ser por ejemplo, yo le llamaría discriminación geográfica, ¿por qué?, los modelos de Tesla por ejemplo, trabajan bien en Estados Unidos pero en México no, por lo baches, por las grietas, por los topes, etc., de hecho hay un Doctor que se llama Uriel, no recuerdo el apellido pero es un Doctor que salió del laboratorio de visión por computadora e hizo una base de datos

de 10 mil imágenes sobre baches y grietas, por ejemplo, ese modelo sí puede ayudar a una cuestión de qué tipo de modelos autónomos pueden funcionar bien, para mí es eso, discriminación geográfica ya que Estados Unidos no tiene ese índice de marginalidad como aquí.
¿Qué aspectos considera que lleva a implementar algoritmos discriminatorios? Pues mira, el caso más conocido, es que las policías de Estados Unidos, por ejemplo, cuando revisan todas las tendencias de los datos de las personas que salen de las cárceles, resulta que los negros o las personas afroamericanas son las que más comenten delitos, entonces cuando la patrulla detecta una placa donde el propietario es una persona afroamericana pues los detiene, o cuando quieren investigar un crimen pues van digamos a buscar a la persona que tiene antecedentes penales.
En su opinión profesional, ¿cuáles deberían de ser las medidas a tomar para evitar este tipo de sesgos? Pues más bien el algoritmo debe tener reglas humanas, no todo se va a resolver creando modelos, o sea se va a tener que agregar algunas capas adicionales para cuando el algoritmo por ejemplo decida, ok esta persona es de Ecatepec, detecto que cometió una infracción y veo que es una persona de Ecatepec con antecedentes penales, aquí tendrías tú que inventar algún tipo de mecanismo humano para que el algoritmo advierta que esa persona, o sea que la tendencia, quién comete delitos son de Ecatepec, pero que tengas cuidado, que te alerte, que no cometas un acto de discriminación y capacitar a las personas también, porque todos estos algoritmos no

> van a reemplazar las funciones humanas, las van a complementar, es como el coche, es una herramienta, pero no quita la función al ser humano de transportarse.

COMENTARIOS PROPIOS COMPLEMENTANDO IDEAS DE LA ENCUESTA

El investigador ha dicho puntos importantes para nuestra investigación, reafirmando ideas que hemos planteado y explicaremos más adelante durante el documento. Hace énfasis en lo complicado que es para un algoritmo aprender, más aún, aprender de las características de las personas cuando estas son el objeto de estudio del modelo de machine learning a implementar, pues existe este factor humano que, si no se toma en consideración, nos remite a una mala praxis, resultando en la segregación de minorías y desigualdad entre individuos, inherente a factores no condicionales y subjetivos.

Uno de los puntos que más nos ha interesado y que contemplamos dentro de nuestros sesgos, es que hay discriminación geográfica, donde se toman en cuenta zonas marginales y se segrega a partir de este índice, esto es muy importante, ya que exilia a diferentes zonas de las posibles oportunidades de crecimiento que puede haber dentro de cada lugar. Desde ahí vemos la importancia del sesgo en todos los ámbitos que ni siquiera tenemos presentes en el día a día, situaciones en donde las personas, un Estado, un país, son exiliados de maneras inimaginables.

Como idea conclusiva, el investigador nos recalca que para que estos sesgos disminuyan al momento de implementar un algoritmo de machine learning, es de suma importancia que el algoritmo cuente con reglas humanas, pues al final son herramientas que no dejarán de lado al

ser humano y esto nos ayuda a complementar nuestro trabajo, es por esta cuestión que las decisiones deben ser al final evaluadas.

ENTREVISTA 2

Nombre del entrevistado: Olga Kolesnikova	**Fecha:** 24 de noviembre del 2022
Cargo: Profesor Investigador	**Lugar:** Centro de Investigación en Computación del IPN

¿A qué se dedica actualmente? Soy profesora investigadora del CIC, mi área es Procesamiento de Lenguaje Natural, entonces aquí imparto asignaturas sobre este tema y también hago investigación.
¿En qué proyectos y cómo ha implementado algoritmos de ML en su trabajo? Pues como tenemos muchos alumnos, yo soy directora de su tesis, entonces hay muchos temas. Los temas ahora que estamos investigando es la detección de la semántica en frases, luego traducción automática neuronal, detección de sentimientos en textos y detección del idioma en los textos donde hay mezcla de idiomas.

¿Cómo describiría en palabras propias la discriminación algorítmica? De hecho, no conozco, o sea en teoría yo leía un como mensaje porque tengo suscripción al portal deepLearning.ia, entonces ahí publican noticias y opiniones de investigadores, y hoy acabo de leer un correo que entonces enviaron donde dicen que sí hay estos algoritmos que son discriminatorios y también tóxicos, chatbots tóxicos, por ejemplo, en teoría lo vi, pero nunca había visto funcionando, ni aplicando en algunas cosas.
¿Ha oído de algún caso de discriminación algorítmica? Si vamos a tener algoritmo de clasificación, si sus clases interpretamos como una clase que vamos a discriminar y otra clase que no, entonces tal vez sería este algoritmo discriminatorio, pero en realidad en práctica, no lo conozco.
¿Qué aspectos considera que lleva a implementar algoritmos discriminatorios? Empezando con colección de datos, entonces hay métodos manuales para recolectar los datos y métodos automáticos entonces, estos métodos hay que revisar por ejemplo, que tipo de población por ejemplo estamos estudiando, si queremos como por ejemplo uno bastante popular, relacionar el sexo de la persona con su nivel de inteligencia, entonces hay estudios en este tipo, pero entonces aquí, esto puede ser sesgado por los tipos tradicionales de actividades que las personas realizan de diferentes sexos, como para una mujer es más frecuente estar en casa en comparación con el hombre, entonces

solamente esta actividad no apoya a desarrollo intelectual de la mujer, entonces por eso si tomamos estos datos nada más mujeres y hombres vamos a medir el nivel de inteligencia puede ser más baja la inteligencia, pero eso no es por sexo, si no por los tipos de actividades tradicionales aquí en la sociedad que están haciendo, eso es ejemplo de sesgo. Entonces aquí se olvida que hay otra variable, otro predictor que tiene su impacto en el nivel de inteligencia que es precisamente el tipo de trabajo que hace la persona, y si no se toma en cuenta esta otra variable entonces vamos a tener resultados sesgados, eso puedo imaginar.
En su opinión profesional, ¿cuáles deberían de ser las medidas a tomar para evitar este tipo de sesgos? Estadísticas e investigaciones de este tipo que mencioné, para tomar decisiones en corte y eso puede ser muy graves porque con base en algunas estadísticas sesgadas van a concluir que esta persona es culpable en esto, van a mandar a cárcel por muchos años y hay casos reales de este tipo, entonces hay que estudiar bien la estadística.

COMENTARIOS PROPIOS COMPLEMENTANDO IDEAS DE LA ENCUESTA

La relevancia de la discriminación algorítmica en su totalidad es un tema que todavía es nuevo en el ámbito, por lo mismo, se están dando casos documentados que muestran cómo los algoritmos tienden a discriminar, el cómo hacer para que esto no suceda es un tema al que apenas se le está dando un acercamiento, objetivo mismo de esta investigación.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Menciona otro punto importante acerca de que la sociedad encapsula actividades para cada sexo, cuando en realidad deberían, como explicaremos más adelante, considerar solamente el conocimiento y la capacidad de cada persona para desarrollar la tarea que se está buscando saciar, no la apariencia o las características que hacen a una persona, ya que, para la vida diaria, no se debería considerar esos aspectos debido al sesgo que generan, daremos ejemplos más adelante que clarifiquen este punto.

Como idea conclusiva podemos notar que hay una creciente visibilidad a estos temas dentro del ámbito científico, ya que hay más investigadores publicando de este tema en mayor cantidad. Actualmente se hacen presentes estos análisis formales o informales publicados en redes sociales que le dan visibilidad a estos temas que recientemente se volvieron parte del foco juvenil, lo que ha causado aún más ruido del que ya existía.

ENTREVISTA 3

Nombre del entrevistado: Amadeo José Argüelles Cruz	**Fecha:** 24 de noviembre del 2022
Cargo: Profesor Investigador	**Lugar:** Centro de Investigación en Computación del IPN

¿A qué se dedica actualmente? Soy profesor investigador del área de ciencias de la computación en el CIC.
¿En qué proyectos y cómo ha implementado algoritmos de ML en su trabajo? Oh bueno, yo tengo proyectos de aprendizaje automático en el reconocimiento de actividades de conducción, estrés en la conducción de vehículos y estoy también trabajando en actividades de reconocimiento de patrones en smart cities, esos son los dos últimos proyectos en los cuales me estoy moviendo y que también tiene que ver con generar resultados de la revisión de sensores que están ubicados en un territorio inteligente, en este caso el territorio es todo Zacatenco, donde estamos midiendo el

censado de contaminantes atmosféricos, movilidad, seguridad, energía y creo que son todos, sí, son todos.
¿Cómo describiría en palabras propias la discriminación algorítmica? Bueno, hablando de sesgos puede haber diferentes tipos de discriminación, pero más bien derivados de que no se hace un análisis desde el inicio sobre qué elementos deben de contemplarse al momento de realizar la implementación de un algoritmo. Y eso parte de la condición de qué datos estoy empleando para poder hacer la implementación de aprendizaje automático, ese es para mí el dato principal. En una etapa previa a la descripción a la implementación del algoritmo debe de existir un conjunto de técnicas que se puedan emplear para poder hacer la selección de características por utilizar sobre los algoritmos de aprendizaje automático, entonces, en esa línea va, si no se contempla, entonces hay una omisión por parte del que implementa los algoritmos de aprendizaje automático en todo el proceso de obtención de resultados, a partir de los datos.
¿Ha oído de algún caso de discriminación algorítmica? Pues hay ejemplos, digamos cuando hay el uso de una aplicación donde no se contemplan personas que son de determinada etnia o rasgos sociales, ahí hay complicaciones porque no le podría yo ofrecer trabajo a una persona con determinada condición étnica, por el simple hecho de que al desarrollar el algoritmo no contemplé que había otro tipo de personas al rededor del planeta Tierra que tiene capacidades iguales o mayores de las que podría tener una etnia social, ahí es donde yo lo vería, la

contratación de personal, ahí hay un ejemplo clásico. No sé, en el uso de aplicaciones que me permitan obtener servicios, y que por mi condición de que vivo en la Gustavo A. Madero, pues no me los dan porque en otra parte del planeta Tierra dicen: “ah espérate, es que tú no lo puedes usar porque eres de México”, entonces ahí se ve ese sesgo, al final de cuentas vivimos en un mundo en donde el entorno global, a mí no me importa de qué lado del planeta vivas, para desarrollar una investigación yo necesito lentes de realidad aumentada, yo requiero de un celular para poder tener la potencia de cómputo para generar las aplicaciones y que funcione, sin embargo, dado que se adquirieron de China, llegaron a México y se quiere echar a andar la aplicación, al momento que tú descargar la aplicación te dice que “no se puede usar para tu país”, ahí está el sesgo, ¿me explico?, y no se indica por ningún lado, no sé por qué, si por su condición de estar del otro lado del planeta y no hablar o pensar como los occidentales, ahí existe también otro sesgo, esos son los que, al menos este que les acabo de decir, es el que me pega a mis actividades laborales, pero hay formas de resolverlo, afortunadamente.

¿Qué aspectos considera que lleva a implementar algoritmos discriminatorios?

Bueno, es que hay que trabajar sobre el aspecto de la Inteligencia Artificial explicable e interpretable, o sea bajo las interpretaciones que empleé que se deben utilizar desarrollar soluciones que incluyan a la Inteligencia Artificial, debo de saber exactamente cómo incluir estos dos elementos , que sea explicable, es decir que yo pueda identificar donde está el sesgo para yo poder darle una solución y una interpretación, de esa manera podríamos desarrollar soluciones que permitan al máximo reducir estos sesgos, que son naturales porque somos humanos no somos máquinas y una máquina va a hacer lo que yo decida como programador, pero siempre

y cuando tengas esos dos elementos importantes como guía para poder desarrollar aplicaciones que no tengan sesgos.
En su opinión profesional, ¿cuáles deberían de ser las medidas a tomar para evitar este tipo de sesgos? Eso va alineado con el desarrollo de un proyecto, en un proyecto se definen claramente las etapas y en una de esas etapas que puede ser la primera y además una que este valorando que se esté cumpliendo el no caer en sesgos, a lo largo del proyecto, pero la primera etapa es clave, decir qué, cómo y para qué, cómo puedo desarrollar una solución que no esté sesgada pues ahí está el cómo, el poder incluir esos elementos a partir del diseño del proyecto en la selección de características y que sea la adecuada porque si le introduzco basura al proceso o al algoritmo, me va a entregar seguramente datos sesgados y eso es lo que no se persigue. Pero para eso está una revisión y una verificación de que el comportamiento del algoritmo o de la aplicación sea correcto.

COMENTARIOS PROPIOS COMPLEMENTANDO IDEAS DE LA ENCUESTA

El investigador, advierte que estos sesgos que derivan a diferentes tipos de discriminación son debidos a problemas en el inicio del proyecto de machine learning, resultados de la falta de análisis sobre qué elementos deben contemplarse al momento de realizar la implementación de un algoritmo. En esta etapa previa a la descripción de la implementación del algoritmo, afirma, deben existir un conjunto de técnicas a emplear para poder hacer la selección de características

por utilizar en el algoritmo, entonces, si no se contempla, hay una omisión por parte de quien implementa los algoritmos.

Además, agrega que en todo proyecto en el que se pretenda incluir inteligencia artificial, se debe trabajar sobre dos aspectos fundamentales de la IA, su explicabilidad e interpretabilidad, lo que nos permitirá identificar sesgos, darles una interpretación y una solución y de esta forma desarrollar soluciones que permitan al máximo reducir estos sesgos. Argüelles afirma, que estos sesgos son naturales debidos a nuestra naturaleza humana y que siempre y cuando se tenga en mente estos dos elementos como guía, se podrán reducir.

Como idea conclusiva podemos tomar diferentes puntos de su opinión, uno de los más importantes es que antes de empezar cada proyecto, tener claro cuáles son los objetivos de este y pensar en el: cómo, para qué y por qué de este. Saber la calidad de los datos que se están agregando, para que el algoritmo no aprenda de “datos basura” y así, por ende, no tener como resultado los datos que discriminan de diferentes formas.

Asimismo, debemos tener en cuenta que los sesgos siempre existirán, no hay forma de eliminarlos en su totalidad, al menos no en la actualidad, ya que como mencionamos y mencionaremos, las personas llevan sus prejuicios a este tipo de prácticas programáticas y no podemos saber específicamente qué características nos lleva a que se generen este tipo de comportamientos.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

ENTREVISTA 4

Nombre del entrevistado: Miguel Jesús Torres Ruiz	**Fecha:** 25 de noviembre de 2022
Cargo: Profesor Investigador	**Lugar:** Centro de Investigación en Computación

¿A qué se dedica actualmente?

Profesor investigador del Centro de Investigación en Computación, líneas de interés son el cómputo urbano, aspectos de Aprendizaje de Máquina enfocados al ambiente geoespacial

¿En qué proyectos y cómo ha implementado algoritmos de ML en su trabajo?

Todo lo relacionado a ciudades inteligentes, todos aquellos fenómenos que son continuos, como el clima o el tráfico, es muy susceptible a que estos enfoques sean utilizados ya que tenemos grandes volúmenes de datos que nos pueden ayudar a caracterizar un fenómeno en particular que aqueja a una ciudad, por ejemplo la contaminación del aire y poder predecir cómo será ese comportamiento en un futuro y a partir de ello tomar medidas, políticas para prevenir o garantizar como se puede solucionar esa problemática, todo enfocado en un ambiente geoespacial.

¿Cómo describiría en palabras propias la discriminación algorítmica?

Está fundamentalmente orientada a todas aquellas herramientas que tienen algún proceso que tiene que ver con la inteligencia artificial. Sin embargo, quienes programa o quienes implementan o diseñan estos algoritmos, al final de cuentas son los humanos, entonces eso puede generar algunos sesgos que pueden perjudicar a un cierto grupo en particular y estos sesgos pueden ser en cuestión de género y esto es al final de cuentas un problema grave, sin embargo este es un tema que ha estado en auge en la sociedad pues el cómo evitar estos sesgos depende mucho dejar de lado prejuicios y que las personas acepten las diferencias entre ellas, esto es algo que una máquina no puede hacer.

¿Ha oído de algún caso de discriminación algorítmica?

En Inglaterra, se generó para una universidad un sistema para admitir estudiantes, buscando automatizar el proceso de admisión pues antes se reunía un grupo de profesores para dar cuenta de quien era admitido y quien no, resultando que el programa intentaba imitar en un 90% el comportamiento de estos profesores y pues en lugar de ayudar a que toda la comunidad de estudiantes tuvieran las misma oportunidades para ingresar, generaba más problemas, porque existían estos sesgos que en la práctica limitaban el acceso a mujeres, personas de diferente etnia o por su tez de color, entonces como resolverlo es complicado, pues debe de haber un grupo diverso y no únicamente en el aspecto técnico, sino también personas preparadas en el aspecto social, para poder hacer que estas herramientas ayuden y no perjudiquen a la sociedad.

¿Qué aspectos considera que lleva a implementar algoritmos discriminatorios?

El hecho de considerar el género es uno de los factores fundamentales, el hecho de pedir el género a una persona ya lo está condicionando, ósea realmente, ¿Importa saber el género? Digo, se supone que tenemos las mismas capacidades y aptitudes para realizar cualquier tipo de trabajo o cualquier actividad en particular, este no debería ser considerado en un sistema tan simple como una encuesta.

Realmente es complicado, lo importante es que la sociedad y las personas que toman decisiones ya están empezando a discutir sobre el tema, como hacer no hay una solución todavía, interesa la etnia, la religión, el género. Todo aquello que es muy subjetivo y puede prestarse a que una persona discrimine a otra no debería de ser considerado para un algoritmo.

Hay algo que se conoce como Feature extraction, con base en los datos se trata de obtener las características para encontrar algún match con un tema en particular, pero yo creo que todo eso que da pie a la subjetividad o a que uno juzgue a alguien no debería de ser considerado. Por ejemplo, en el CIC no manejamos nombres cuando tenemos aspirantes, entonces tienen etiquetas los aspirantes, todos tienen etiqueta y no sabemos si esa etiqueta es hombre o mujer, tampoco de donde egresas porque muchos prefieren egresados del politécnico, para que aspiren a entrar aquí que de algún otra escuela y eso al final de cuentas es discriminación, pues a lo que te tienes que avocar es a una prueba a aprobar un examen y si eres totalmente capaz de aprobar ese examen tienes todo para ingresar, o también el promedio, entonces al final de cuentas si quisiéramos meter todas esas cosas en un algoritmo seria discriminatorio, entonces yo creo que todas esas cosas no deberían considerarse, que en términos de eficiencia

del algoritmo si impacta, porque se busca tener a los mejores estudiantes al menos en este ejemplo, pero todos deberían tener las misma oportunidades.

En su opinión profesional, ¿cuáles deberían de ser las medidas a tomar para evitar este tipo de sesgos?

Repito, no considerar las características de los individuos como tal y todos aquellos elementos relacionados que permitan juzgar a una persona y no creo que estos elementos o características que describen a un individuo no deben formalizar reglas, entonces todo aquello que condicione a priorizar a una persona o seleccionar a un grupo sobre de otro no debe ser una regla o parámetro para formar parte de un algoritmo.

COMENTARIOS PROPIOS COMPLEMENTANDO IDEAS DE LA ENCUESTA

En la implementación de algoritmos de machine learning, nos comenta que se pueden generar algunos sesgos que pueden perjudicar a un grupo en particular. Es importante destacar su opinión acerca de los aspectos que debemos tener en cuenta para hacer un análisis, muchas veces, como menciona él, no es necesario tener en cuenta el género, por ejemplo, podemos preguntarnos, ¿en qué diferencia que sea una mujer a un hombre para realizar la actividad de la cual se pidió ese campo?, ¿el sexo cambia algo conforme al conocimiento de la persona?, aplica de igual forma para personas que son discriminadas racialmente.

Como idea conclusiva, el investigador realza la importancia de saber qué características seleccionar en el conjunto de datos que entrenará al algoritmo, destacando, que en su opinión profesional todas aquellas características inherentes a un individuo y aquellos elementos

relacionados que permitan juzgar a una persona lleguen a formalizar reglas que prioricen a un grupo o individuo sobre otros.

Investigación

En esta sección se buscará dar un porqué al existir de los algoritmos discriminatorios, usando las entrevistas realizadas y analizando situaciones donde se han presentado este tipo de algoritmos con la finalidad de explicar qué factores afectan a su creación.

Capítulo 1: Como la inteligencia artificial llega al sesgo

Un algoritmo solo puede ser tan bueno como los datos que le damos. Si un algoritmo está minando en una sección de nuestros datasets que, por alguna razón, es poco representativa de la población, esto producirá un resultado no representativo.

Estas variaciones en la calidad y representatividad de los datos pueden estar correlacionado con la pertenencia a clases y puede impactar negativamente a grupos que han sufrido, a lo largo de la historia, de desventaja cuando usamos estos datos para realizar decisiones sobre miembros de estos grupos.

La solución a este problema usualmente se resolvería obteniendo más datos, ya que, cuando una muestra es lo suficientemente grande esta va a parecerse a la población. Sin embargo, en este contexto obtener más datos la mayoría de las veces falla no logra resolver el problema, esto o porque el algoritmo ya usa una base de datos sesgada o porque se siguen obteniendo muestras sesgadas.

Los datos de entrenamiento para un algoritmo de machine learning no solo pueden ser sesgados o incompletos, también estos pueden reflejar discriminación posterior. Por ejemplo, si entrenamos a un algoritmo con datos de aplicantes aceptados a una universidad donde antes

este proceso lo hacían profesores, esos datos estarán tan sesgados como los profesores en sí y el algoritmo replicará los mismos sesgos que hacían los profesores.

1. Datos sesgados

Barocas y Selbst destacan seis razones por las que la inteligencia artificial llega a la discriminación por medio de los datos. Los problemas se relacionan con cómo definimos la variable objetivo y las etiquetas de clase, etiquetar los datos de entrenamiento, la recolección de los datos, selección de características, proxies y discriminación intencional [20].

a) Definir La Variable Objetivo Y Las Etiquetas De Clase

Al definir la variable objetivo tenemos que preguntarnos cómo podemos calcularla, por ejemplo, a la hora de crear algoritmos de despidos. ¿Cómo definimos a un mal trabajador? Dependiendo de cuales etiquetas de clase usemos estas pueden tener un impacto negativo en clases sociales bajas. Digamos que tomamos como ejemplo si alguien es puntual. La mayoría de las veces la población pobre no vive cerca de sus trabajos y estos realizan trayectos de largos tiempos para llegar a trabajar, también tomemos en cuenta que usan el transporte público, este no siempre puede ser de confianza, por lo tanto, tienen las de perder si el algoritmo favorece la puntualidad.

b) Etiquetado De Los Datos De Entrenamiento

Si entrenamos al algoritmo con datos de entrenamiento ya sesgados, estos datos vienen así ya que representan decisiones de personas que optan por desfavorecer minorías, el algoritmo recrea este sesgo. Esto pasó en Amazon en 2014, entrenaron su algoritmo de contratación con 10 años de datos de contratación, lo que pasó fue que los contratadores preferían aplicantes hombres, esté sesgo lo recreó el algoritmo a la hora de darle estos datos para aprender.

c) Recolección De Datos

A la hora de hacer la recolección de datos también puede ser sesgado. Un caso de esto es la recolección de datos sobre crimen, si la policía ha arrestado más a grupos étnicos existe una sobrerrepresentación y el algoritmo aprenderá que son más propensos a cometer crímenes.

d) Selección De Características

Al utilizar los algoritmos, analizar todas las variables para realizar el cálculo puede ser muy costoso, entonces tenemos que simplificar las características que le damos y concentrarnos en cuales son las que mejor caracterizan a lo que queremos obtener. Esto a veces puede introducir sesgo contra las minorías. Por ejemplo, digamos que las empresas prefieren candidatos que vengan de universidades privadas, esto desfavorece a grupos étnicos ya que estos son menos probables que puedan estudiar en una universidad privada ya que son bastante costosas.

e) Proxies

Los proxies son características que pueden llevar al algoritmo a características protegidas de la población. Los algoritmos no pueden utilizar la raza para determinar un resultado, ya que sería más susceptible a ser más discriminatorio, pero, por ejemplo, los bancos que hacen uso de algoritmos de machine learning emplean el código postal y supongamos que este tiene correlación con la raza del individuo esto beneficiaría a algunos, pero perjudicaría a otros basados en su raza indirectamente.
Barocas y Selbst explican que “el problema surge de lo que los investigadores llaman "codificaciones redundantes", casos en los que la pertenencia a una clase protegida resulta ser codificado en otros datos. Esto ocurre cuando una pieza particular de datos o ciertos valores

para esa pieza de datos está altamente correlacionada con la membresía en áreas protegidas específicas clases"[20].

f) Discriminación Intencional

Y, por último, también puede que esta discriminación sea intencional. Supongamos que la compañía quiere despedir a mujeres embarazadas, esto claro no debería ser posible ya que sería un caso de discriminación, pero las empresas usan una justificación para hacerlo. Esto ya ha ocurrido y es el caso de la compañía Amazon que mediante un algoritmo despedían a los trabajadores menos eficientes, esto sin tomar en cuenta claro en las dificultades que, en este caso, puede tener una mujer embarazada. Esto pudo no haber sido intencional, pero no significa que algo así no pudiera pasar.

Estas son las seis razones por las cuales un algoritmo de machine learning puede llegar a ser discriminatorio, desde los datos que le damos, y nos dicen que tenemos que ser cuidadosos a la hora de trabajar con estos. Tenemos que tomar en cuenta las implicaciones que tiene utilizar estos algoritmos y cómo estos pueden agrandar las brechas de desigualdad.

2. Análisis previo a la implementación

Ya vimos cómo se sesgan los algoritmos mediante los datos. ¿Qué podemos hacer con eso? Vamos a analizar qué podemos hacer antes y durante el proceso de entrenamiento del algoritmo para reducir la discriminación algorítmica.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Primero al definir la variable objetivo tenemos que preguntarnos si tenemos la cantidad y calidad de datos, como nos mencionó el Dr. Argüelles, la discriminación en los algoritmos es derivada a un nulo análisis, previo a la implementación, sobre qué elementos deben de contemplarse y como esta primera etapa es clave para definir el desarrollo de una solución no sesgada.

Entonces previo a un proyecto, donde trabajemos con datos personales e implementemos un algoritmo de machine learning, necesitamos ver cómo vamos a encontrar la solución, aquí la parte de definir la variable objetivo, tener en nuestra mente que estos datos pueden tener sesgos, analizarlos antes de que hagamos la implementación viendo si estos son una representación buena de nuestra población, cuestionarnos cómo fue la recolección de estos datos y si esta recolección pudo haber tenido sesgos y finalmente con ese conocimiento saber qué características seleccionar para tener un algoritmo con sesgo reducido.

Aunque, solo podemos hacer lo posible con la información que nos dan y lo más que podemos hacer sería bloquear información, quitarles a los datos que tenemos los que nosotros creamos sensibles, más que quitarlos no considerarlos para el entrenamiento porque al final estos pueden ser útiles para analizar estadísticamente los resultados para notar si existe un sesgo.

Esto viene con un costo, bloquear información puede ser contraproducente, claramente cosas como raza y género no, pero otras características que son proxies puede ser mala idea bloquearlos.

3. Problemas al bloquear proxies

Para aplicar una política de información a algoritmos, es crucial identificar qué información es proxy para otras. El bloquear los proxies puede ser la clave para evitar la discriminación.

Al proponer esta idea surgen dos problemas, el primero es que en realidad no sabemos qué información es la que está actuando como proxy, y aunque lo supiéramos, es muy difícil bloquear todos los proxies. El segundo problema es que si pudiéramos bloquear todos los proxies esto puede ser contraproducente porque estos pueden contener información valiosa.

Con el primer problema tenemos un caso de que toda la información sea proxy y esta afecte poco al resultado, entonces para evitar la discriminación tendríamos que deshacernos de la mayoría de la información con la que estamos trabajando, ya que si nos deshacemos de algunas esto haría un cambio pequeño al resultado.

Supongamos que podemos bloquear todos los proxies y así tener un algoritmo que realice cálculos neutrales, esto llegaría a un punto donde no tendríamos la información suficiente para determinar un resultado óptimo y nos quedamos con algo que tendría a la aleatoriedad como parte del algoritmo.

El segundo problema es que, cuando uno quita cualquier información, sin querer puede quitar información relevante para la toma de decisiones. Algo que sea un proxy para llegar a datos protegidos, también puede ser un proxy para información útil de gran valor.

Si a veces bloquear información es contraproducente. ¿Qué otra cosa podemos hacer para reducir la desigualdad que producen los algoritmos discriminantes?

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Lo lógico sería conseguir más información, pero no es tan fácil porque puede que en la recolección de datos se obtengan los sesgos entonces conseguir más datos puede no hacer nada. Registros individuales pueden sufrir de problemas de calidad dado por datos parciales o hasta datos incorrectos. El dataset puede tener problemas de calidad a mayores proporciones para grupos minoritarios comparado con otros grupos o puede ser no representativo para la población en general. Entonces necesitamos datos de buena calidad para que estos sean representativos de la población.

Sin embargo, mientras que bloquear información de categorías protegidas es inútil, dándole forma a la información que incluye categorías protegidas puede ser efectivo en torno a eliminar sesgos de los datos de entrenamiento que se les dan a los modelos que hacen decisiones, sucesivamente, eliminando la discriminación de esos modelos. Podemos editar los datos de entrenamiento para parecerse a un mundo más igualitario en el que deberíamos vivir.

Capítulo 2: Poca representación en algoritmos discriminatorios

Otra forma en la que se da la discriminación algorítmica ocurre es mediante la subrepresentación de las minorías. Esto puede ser mediante los datos, como ejemplo datos de postulantes aceptados en una compañía por lo cual los algoritmos rechazan a estos, o por la subrepresentación de minoría con puestos de machine learning.

Una de las causas más comunes del sesgo en los algoritmos de machine learning es que en los datos de entrenamiento faltan muestras de grupos poco representados. Por esto los asistentes personales a veces no entienden acentos o algoritmos de reconocimiento de imágenes etiquetan imágenes de personas negras como gorilas. Por esto es bastante importante que nos aseguremos que nuestros datos de entrenamiento tengan representación de todos los grupos poco representados.

1. Poca representación en los datos de minorías

Se han obtenido estadísticas en donde se reitera que consecuentemente las mujeres y las minorías se ven afectadas dentro de muchas situaciones de vida. La mayor parte del mundo se ha segado a que solo existen mayoritariamente los hombres blancos, dejando de lado a todas las demás etnias.

Cuando un país crea algo que tenga que tener Inteligencia Artificial de por medio, posiblemente ya vaya sesgado para personas que viven del otro lado del mundo con respecto a ellos, ya que no están considerando incluir su lenguaje, por ejemplo, retomando la entrevista del profesor investigador Zagal Eswart, él nos dice desde su experiencia y conocimiento que Tesla es una

de estas empresas en donde para México no funciona su IA, debido a que no considera que tenemos diferentes condiciones a las de su país, sin embargo, aun así, vende su coche dentro de nuestro país, lo que nos haría pensar que también consideró las condiciones en las que se vive en diferentes zonas. Otro ejemplo, lo vemos con la entrevista del profesor investigador José Amadeo, quién nos dice que para su trabajo hay veces que también tiene impedimentos debido a que el país de dónde vienen los productos pedidos para realizar el mismo, no existen aquí. Debido a la misma situación de que no consideran venta para ciertos países o alguna aplicación no funciona aquí, no puedes cambiar el idioma porque solo pusieron el de ellos, entre muchos ejemplos más que nos dejan claro que la poca representación está siendo ya un problema desde antes, sin embargo, apenas se empezó a alzar la voz porque cada vez hay más personas que luchan por la igualdad de condiciones, sean vistos por su intelecto, no por su aspecto.

2. Falta de diversidad

Parte de todo este problema, es que no hay diversidad en cuanto al personal de programación, el promedio es: blanco, varón, técnico, con educación formal, de habla inglesa; este es el contenido del que aprenden los algoritmos. Sucede lo mismo con los datos, los datos que se utilizan para el comportamiento humano, el entrenamiento de las IA se basa en las muestras *WEIRD*, acrónimo de White (Blanco), Educated (Educado), Industrialized (industrializado), Rich (rico) and Democratic (democrático). [21]

Para un campo que supuestamente está remodelando la sociedad como lo son las ciencias de la computación, resulta preocupante observar la homogeneidad de los investigadores dentro del ámbito de la Inteligencia Artificial y es que esta poca diversidad lleva a la creación de sistemas sesgados que representan de manera pobre a minorías. La inteligencia artificial se ha vuelto una parte fundamental de nuestra vida, pero ¿Qué hacemos si toda esta tecnología masiva está

sesgada involuntariamente? y ¿qué hacemos si este campo de estudio es investigado por un sector definido de la población que no la representa?

La principal razón por la que necesitamos diversidad, tanto en los conjuntos de datos como en los grupos de investigadores es porque se necesita de personas que tengan el sentido social de cómo son las cosas. Desde un punto de vista técnico, hay distintos acercamientos para reducir los sesgos. Uno es diversificar el conjunto de datos y tener anotaciones sobre datos relevantes del mismo, una vez entrenado el modelo, probar que tan bien trabaja con diferentes subgrupos. Incluso siguiendo esto, no podemos asegurar que no exista un sesgo, puesto que no se puede tener un conjunto de datos que defina a la perfección a todas las personas y un modelo que no generaliza, no es un buen modelo.

Se plantea que en realidad los algoritmos no son los nuevos campos de discriminación ya que ellos se programan por el humano para reproducir virtualmente la realidad y con ellos sus profundas desigualdades sociales. Afirman que esto es debido a que los programadores, son mayoritariamente hombres, que como colectivo se han preocupado muy poco por los problemas sociales que existen dentro de este ámbito, por lo que no es casualidad que las escasas iniciativas que existen de incorporar valor social al diseño de la inteligencia artificial provengan de mujeres. Afortunadamente, esto es reversible, la programación podría impregnarse de otros valores humanos más próximos a la equidad y a la igualdad. [22]

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Capítulo 3: Sesgo producido por la gente

Algo que podemos concluir de los capítulos pasados es que, en sí, los algoritmos no son los que, por así decirlo, generan los sesgos sino la gente que los implementa. Esto puede ser por lo ya discutido en los capítulos anteriores que puede ser porque no realizaron un proceso previo a la implementación además de que los datos están llenos de prejuicios.

Cada vez más evidencia sugiere que los prejuicios humanos se han integrado en estas herramientas porque los modelos de aprendizaje automático se entrenan con datos sesgados. Lejos de evitar el racismo, simplemente pueden ser mejores para ocultarlo. Muchos críticos ahora ven estas herramientas como una forma de lavado de tecnología, donde una apariencia de objetividad cubre los mecanismos que perpetúan las desigualdades en la sociedad. El reto consiste en incorporar algoritmos, aprendizaje automatizado y la inteligencia artificial para evitar sesgos humanos, no para reproducirlos. Infortunadamente, los casos de algoritmos los cuales han contribuido a empeorar los procesos, gracias a esto, se han levantado las voces para abogar a favor de la transparencia algorítmica.

No debe sorprendernos que los algoritmos muestran prejuicios sociales. La IA no es magia, los humanos somos propensos a una interpretación sesgada de la realidad. Esta discriminación pasa desapercibida porque la mayoría de la gente no podemos reconocer nuestros propios sesgos, esto se conoce como *punto ciego de los sesgos*. En la IA nuestros prejuicios individuales se combinan con los de otras personas en una escala masiva y se amplifican, por ello la discriminación se enmascara y se complica detectar sesgos. [23]

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

5C

Propuesta de Normas para un Protocolo

Introducción

Con la emergente aplicación de algoritmos de machine learning en diversas industrias y en nuestro día a día, han surgido casos donde la implementación de estos algoritmos ha resultado en situaciones de desigualdad para ciertos grupos minoritarios y sectores de la sociedad.

Es por ello, que se plantean las siguientes propuestas para que formen parte de un protocolo que sirva como guía en todo trabajo o proyecto donde se pretenda implementar un algoritmo de machine learning, con la finalidad de disminuir los sesgos y de esta forma reducir los casos donde las minorías son segregadas y poco representadas.

Objetivo

El presente protocolo busca instruir a realizar un proceso de análisis previo a la implementación de un algoritmo de machine learning dentro de cualquier proyecto o trabajo de inteligencia artificial. Buscando que los resultados que se obtengan representen de una forma más fidedigna a los grupos que suelen ser segregados.

Con esta finalidad en mente, se postulan algunas propuestas a tomar en cuenta antes de empezar un proyecto de machine learning, pues, su consideración e implementación llevarán a disminuir sesgos de índole racial en cualquier proyecto.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Objetivos específicos

Objetivo 1: Encontrar y describir las complicaciones que existen para detectar los sesgos raciales de los algoritmos.

Objetivo 2: Identificar las normas que consideramos eficaces que logran hacer un correcto proceso de limpieza de datos y que ayudan a disminuir el sesgo racial.

Objetivo 3: Validar la eticidad detrás de cualquier proyecto de machine learning.

Objetivo 4: Hacer un correcto uso tanto de los datos recolectados como de los algoritmos de machine learning por implementar.

Objetivo 5: Enfatizar la relevancia de la planeación dentro de un proyecto de machine learning y la calidad de los datos.

Objetivo 6: Identificar algunas características que puedan hacer que los datos presenten sesgos, debido a la subjetividad y naturaleza de las propias características seleccionadas.

Propuesta

Del análisis fundamentado en la presente investigación se concluyeron algunos puntos que los investigadores consideramos clave para reducir sesgos raciales al momento de implementar algoritmos de machine learning.

- Pensar en explicabilidad e interpretabilidad

Con este punto nos referimos a que todo proyecto de inteligencia artificial debe cumplir con estos dos puntos.

Haciendo referencia a que todos los resultados que dan como solución las inteligencias artificiales deben buscar ser entendidos por expertos humanos en el objeto de estudio. Entendiendo que muchas veces los algoritmos que se implementan funcionan como cajas negras donde quien lo implementó no sabe el porqué de la decisión que tomó el modelo de machine learning, esto trae consigo un reto técnico de interpretabilidad, sin embargo, se incita al lector a buscar el equilibrio entre la interpretabilidad y la integridad del sistema desarrollado.

Tener esto en mente al desarrollar una inteligencia artificial, nos permite identificar fácilmente sesgos que se puedan haber generado durante el proyecto y darles solución.

- Pensar en la ética del proyecto
    - El impacto que pueden tener
      Se tiene que internalizar la idea de qué consecuencias tendrá implementar un algoritmo teniendo en cuenta si este puede agrandar las brechas de desigualdad ya existentes.
    - Prejuicios que se tienen como ser humano
      Recordemos que los sesgos se generan por las creencias de los que lo programan, ya que los algoritmos son programados por nosotros, aprenden de lo que nosotros les proporcionamos, si nosotros les ofrecemos discriminación racial, de

género, geográfica, entre muchas otras, eso es lo que nos dará como resultado, un algoritmo sesgado por nosotros mismos.

- Datos
    - Calidad

      Se necesita una calidad de datos que sea diversa y por ende representativa de la población. Cuando los datos no tienen calidad ya sea por su método de recolección, o porque la muestra está desbalanceada, hay técnicas que permiten trabajar mejorar los datos como el sobre muestreo y el submuestreo, pero si el conjunto de datos no representa al problema a atacar de forma prudente es válido cuestionarse si vale la pena usarlos.
    - De donde provienen

      Se tiene que tomar en cuenta si el proceso de recolección de datos es justo o está sesgado, ya que si entrenamos un algoritmo con datos sesgados este los va a replicar. Por consiguiente, las inteligencias artificiales recopilan nuestros prejuicios individuales se combinan con los de otras personas en una escala masiva y se amplifican, por ello la discriminación se enmascara y se complica detectar sesgos.

- Selección de características
    - En este punto, tenemos que considerar las cualidades intelectuales de las personas, como lo pueden ser: agilidad mental para resolver problemas, raciocinio, análisis dentro de ciertos escenarios, entre otros más. Hay que

recalcar que se debe entender el objetivo de la investigación desde un principio, incluso antes de partir en elegir características, para así hacer una selección adecuada de éstas que se utilizaran dentro del proyecto. Preguntarnos si realmente saber el sexo de la persona nos apoyará en tomar una decisión argumentada sobre lo que puede hacer o no una persona, cuando lo que se necesita conocer es, como ya se dijo, la capacidad intelectual de la persona.

- Grupos de investigación diversos
    - Tenemos que considerar ampliar la contratación en los campos científicos y programadores, ya que es el área en donde más sesgos se generan. Tenemos que tomar en cuenta que no solo deben ser hombres blancos de habla inglesa, si no tener en cuenta, diferentes etnias, razas, idiomas, etc. Para que así los datos tengan la diversidad necesaria, con personas que no tengan un sesgo racial sobre otros.

Conclusión

Usualmente, no se pone mucha atención al cómo se recolectan, procesan y organizan los datos, llevando así a que algunos grupos sean representados en gran medida y otros son infrarrepresentados.

Estos sistemas de machine learning están siendo utilizados para tomar decisiones que cambian vidas. Estas decisiones pueden llegar a perjudicar los derechos humanos, usualmente de las personas más vulnerables de la sociedad.

Bajo un buen uso y diseño, los sistemas de machine learning, pueden ayudar a eliminar algunos sesgos en la toma de decisiones que afectan a la sociedad. Sin embargo, también es posible para estos sistemas reforzar el sesgo sistemático y la discriminación, eludiendo la garantía de dignidad humana.

Los resultados discriminatorios no sólo violan los derechos humanos, sino que también demeritan la confianza pública en los algoritmos de machine learning, dificultando el desarrollo del campo y mitigando su potencial tanto social como económico. Es por ello qué se necesitan soluciones sistemáticas para atacar el problema de discriminación algorítmica.

Afortunadamente, con esta investigación se ha podido concluir ciertas acciones que ayudan a disminuir los sesgos que provocan la discriminación algorítmica con la implementación de algoritmos de machine learning.

Instituto Politécnico Nacional

Escuela Superior de Cómputo

De Luna Ocampo Yanina, Ramírez Méndez Kevin, Sainz Takata Juan Pablo Minoru

Referencias

1. Arnold, David, Will Dobbie, and Peter Hull. 2021. "Measuring Racial Discrimination in Algorithms." *AEA Papers and Proceedings*, 111: 49-54.
2. How Some Algorithm Lending Programs Discriminate Against Minorities. (2018, 24 noviembre). Npr. Recuperado 24 de agosto de 2022, de https://choice.npr.org/index.html?origin=https://www.npr.org/2018/11/24/670513608/how-some-algorithm-lending-programs-discriminate-against-minorities
3. K. Chmielinski, Data Nutrition Project, https://datanutrition.org/#section-solution-research
4. *García, R. (2006). Epistemología y Teoría del Conocimiento Epistemology and. Redalyc, 2(salud colectiva), 113–122.*
5. Glaser B. Conceptualization: On theory and theorizing using grounded theory. Int J Qualit Methods. 2002:1
6. R. Burcio, *Discriminacion Racial a La Mexicana*, CONAPRED, https://www.conapred.org.mx/index.php?contenido=noticias&id=3766&id_opcion=&op=448 (Accedido el 26 de agosto de 2022).
7. *Una de cada cinco personas de 18 años y más, declaró haber sido discriminada, Encuesta Nacional sobre Discriminación (ENADIS), 2017*
8. P. Rivas Vallejo, *Discriminación algorítmica: detección, prevención y tutela*. Madrid, 2021.
9. "algoritmo | Diccionario de la lengua española". «Diccionario de la lengua española» - Edición del Tricentenario. https://dle.rae.es/algoritmo(accedido el 24 de noviembre de 2022).
10. T. Alameda. "Te contamos qué es el 'machine learning' y cómo funciona". BBVA NOTICIAS. https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/(accedido el 26 de noviembre de 2022).

11. M. Gilman, “Poverty lawgorithms,” *Data & Society*. [Online]. Available: https://datasociety.net/library/poverty-lawgorithms/. [Accessed: 20-Nov-2022].
12. E. Sheng, K.-W. Chang, P. Natarajan y N. Peng. "The Woman Worked as a Babysitter: On Biases in Language Generation". ACL Anthology - ACL Anthology. https://aclanthology.org/D19-1339.pdf (accedido el 24 de noviembre de 2022).
13. A. Caliskan-Islam, J. J. Bryson y A. Narayanan. "Semantics derived automatically from language corpora necessarily contain human biases". Computer Science Department at Princeton University. https://www.cs.princeton.edu/~arvindn/publications/language-bias.pdf (accedido el 24 de noviembre de 2022).
14. J. Kasperkevic. "Google says sorry for racist auto-tag in photo app". the Guardian. https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app (accedido el 24 de noviembre de 2022).
15. Z. Obermeyer, B. Powers, C. Vogeli y S. Mullainathan. "Dissecting racial bias in an algorithm used to manage the health of populations". Federal Trade Commission | Protecting America's Consumers. https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-ziad_obermeyer.pdf (accedido el 24 de noviembre de 2022).
16. "AlgoRace". AlgoRace. https://algorace.org(accedido el 26 de noviembre de 2022).
17. M. Velasco. "Cómo los algoritmos te discriminan por origen racial y por género". ElHuffPost. https://www.huffingtonpost.es/entry/como-los-algoritmos-te-discriminan-por-origen-racial-y-por-genero_es_62989efae4b07aa938968a36(accedido el 26 de noviembre de 2022).
18. B. Heinrichs. “Discrimination in the age of Artificial Intelligence”, AI & SOCIETY, 2021

19. Ignacio N. Cofone, Algorithmic Discrimination Is an Information Problem, 70 Hastings L.J. 1389 (2019)
20. S. Barocas y A. D. Selbst, "Big Data's Disparate Impact", *SSRN Electronic Journal*, 2016. Accedido el 22 de noviembre de 2022. [En línea]. Disponible: https://doi.org/10.2139/ssrn.2477899
21. S. Dobrin y S. van den Heever. "Putting Diversity to Work in Data Science - THINK Blog". THINK Blog. https://www.ibm.com/blogs/think/2019/12/ibm-diversity-in-data-science/ (accedido el 24 de noviembre de 2022).
22. M. Velasco. "Cómo los algoritmos te discriminan por origen racial y por género". ElHuffPost. https://www.huffingtonpost.es/entry/como-los-algoritmos-te-discriminan-por-origen-racial-y-por-genero_es_62989efae4b07aa938968a36(accedido el 26 de noviembre de 2022).
23. E. Pronin, D. Y. Lin y L. Ross. "The Bias Blind Spot: Perceptions of Bias in Self Versus Others". Psykologi, narsissisme,psykopati og sakkyndige. http://www.sakkyndig.com/psykologi/artvit/pronin2015.pdf (accedido el 24 de noviembre de 2022).