

# 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. (2022)

De Luna O. Yanina, *ESCOM, IPN*, Sainz T. Juan, *ESCOM, IPN*, Ramírez M. Kevin, *ESCOM, IPN*

**Abstract—** In this article we investigated that with the pass of time, the use of Artificial Intelligence models has grown all around the world, consequently, inequality gaps towards minority groups have been magnified by the usage of biased algorithms as proposed by different authors. Furthermore, we must assume the existing debate concerning to why does racial discrimination exists, and we will present our own definition about this concept. We explain that an important cause to be considered is the lack of diversity within the programmers in the tech industry who are those in charge of implementing these algorithms, not only because the average programmer is a white male but also because of the gathered data which does not properly represent minorities. Hence, increases bias within the data, leading the algorithms to reflect the already existent discrimination in our society. Finally, we sustain the claim that this lack of diversity increases the racial problematic that is already existent inside multiple training data for an algorithm with a social end, as consequence algorithms learn from this data and generate diverse problems in our current society. That is why with the current implementations of Artificial Intelligence in Mexico we do not wish that these problems that we can observe be replicated in Mexican society so we will propose norms for a protocol that has as a goal to minimize the bias that these algorithms have.

**Index Terms—**algorithm, bias, inequality, minority, models, racial discrimination.

## I. INTRODUCCION

La argumentación presentada con este trabajo procede en tres pasos: la primera parte es acerca del racismo y los algoritmos racistas, donde se va a contextualizar acerca de que lo que se tiene presente en la sociedad con respecto a estos temas.

Se argumenta que la discriminación en sí es inapropiada para la justicia ética. La noción de la discriminación no significa que sea totalmente inútil, más bien

que uno mismo debe poner atención a si el concepto es aplicado correctamente.

Siguiendo con esto, en la segunda parte afirmamos que con la Inteligencia Artificial se pueden agravar los problemas de discriminación, pero de una manera diferente de lo que la mayoría de los críticos suponen. Con la implementación de esta Inteligencia Artificial para tomar decisiones, se vuelve mucho más difícil poder determinar si ciertas instancias están en desacuerdo con una ponderación establecida de las preocupaciones éticas pertinentes. Se debe a su opacidad epistémica que ésta amenaza con socavar exactamente el tipo de deliberación moral que es esencial para llegar a un entendimiento en vista de lo que debe contar como discriminación. Sin embargo, los algoritmos también pueden ayudar a detectar formas ocultas de discriminación.

Por último, nuestro tercer paso es entender por qué afirmamos el que la poca representación dentro de estos temas y la discriminación de las minorías afecta de sobre manera la forma en la que los algoritmos se convierten en algoritmos raciales, aprendiendo de los datos que se les brindan, estos datos ciertamente ya tienen impuestas las creencias de las personas que los están manipulando. El que no haya diversidad, hace que se impongan las creencias de solo un grupo de personas dentro de estos datos, creando estos sesgos ya mencionados.

## II. CONTEXTUALIZACIÓN

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. 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.

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 [4].

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. 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 [5].

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.

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, teniendo de esta forma su auge en las últimas décadas en usos desde el arte, hasta la medicina. 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.

### III. METODOLOGÍA

Durante la investigación se escogió una forma de abordar el tema de estudio, en esta sección se pretende esclarecer dicha forma, desde el sustento teórico, corrientes filosóficas

abordadas para la obtención del conocimiento remarcadas en el enfoque epistemológico de la investigación, delimitando la población objeto de estudio y las técnicas de recopilación y análisis de datos realizadas durante el trabajo. A forma de prefacio, se conceptualizan las variables que fueron empleadas en el estudio.

- Independiente: Análisis de algoritmos discriminatorios.
  - Entendemos por el análisis de algoritmos discriminatorios a la observación de casos aplicados y prácticos donde la implementación de algoritmos de aprendizaje de máquina donde por algún motivo el resultado dio pie a situaciones de discriminación y desigualdad en la sociedad.
- Dependiente: Propuesta de normas para un protocolo que disminuya la discriminación racial algorítmica
  - Bajo el entendido de esta variable, se interpreta que del análisis realizado a los diferentes algoritmos discriminatorios se pretende encontrar factores en común indicadores de un patrón descriptible y explicable de las razones por las cuales los algoritmos discriminan, para con ello fundamentar propuestas de un protocolo donde se procure disminuir las situaciones de discriminación algorítmica en la implementación de algoritmos aprendizaje de máquina.

El enfoque epistemológico que toma la investigación está dado por la corriente filosófica del empirismo lógico, debido a que nuestro análisis sienta sus bases directamente en evidencias y eventos observables, como lo son la implementación de algoritmos discriminatorios, esto a su vez nos permite hacer uso de metodo inductivo, cuya característica principal es partir de lo particular a lo general, partiendo en este caso de las observaciones donde algoritmos generan desigualdad los casos particulares de los cuales se pretende obtener patrones en común que nos lleven a definir y explicar el porqué de estos eventos.

Esta investigación toma el paradigma cualitativo, permitiendo contextualizar y tener una mayor inmersión en el fenómeno objeto de estudio, a partir del análisis de correlaciones de causa y efecto entre las variables. Asimismo, esta investigación no se fundamenta en la estadística y no pretende seguir una secuencia lineal, es una investigación cuyo propósito es buscar, analizar e interpretar información documental y la recopilada con los instrumentos utilizados.

Afín 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 microsociológica centrándose en la estructura social, los sistemas sociales y la población a gran escala, pues el análisis involucra a como estos algoritmos afectan en general a la sociedad.

De lo anterior, se puede definir como población objeto de estudio a los casos de implementación de algoritmos de aprendizaje de máquina que han dado lugar a situaciones de discriminación y desigualdad en la sociedad alrededor del mundo, además, como complemento se entrevistó a un grupo diverso de expertos en el desarrollo de algoritmos de aprendizaje de máquina.

En la investigación se usó como técnica de recopilación de datos la entrevista a expertos en el tema con la finalidad de esclarecer la ocurrencia del fenómeno estudiado y tener opiniones validas y variadas con respecto a las causas de ocurrencia de discriminación algorítmica. El instrumento usado fue validado por el ingeniero de aprendizaje de máquina y desarrollo de software José Martín Ávila Martell, haciendo énfasis que el uso de la información recolectada será meramente divulgativo e informativo. El siguiente es el guion de la entrevista usado.

Nombre del entrevistado:	Fecha:
Cargo:	Lugar:
¿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?	

*Ilustración 1. Guion de la Entrevista*

#### IV. DESARROLLO

En esta sección ahondaremos minuciosamente en las entrevistas realizadas a expertos investigadores en Machine Learning, recapitulando las ideas primordiales que obtenidas que llevaron a definir propuestas para a favor de reducir la implementación de algoritmos discriminatorios.

Reafirmando nuestra hipótesis, el investigador Zagal Flores Esward 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. Recalcando, además, la relevancia del proceso de ETL durante el proyecto de Machine Learning y también para que estos sesgos que llevan a discriminación 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.

Por otra parte, el investigador Amadeo José Argüelles Cruz, 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.

En adición, el investigador Miguel Jesús Torres Ruiz realza la importancia de saber que 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.

#### III. RESULTADOS

Las siguientes son las propuestas a tener en consideración al momento de desarrollar un proyecto de Machine Learning.

- Pensar en explicabilidad e interpretabilidad. Todos los resultados que dan como solución las inteligencias artificiales deben buscar ser entendidos por expertos humanos en el objeto de estudio.
- Pensar en la ética del proyecto

- El impacto que pueda tener  
Se tiene que pensar en la idea de qué consecuencias tendrá implementar un algoritmo.
- Prejuicios que se tienen como ser humano  
Los sesgos se generan por las creencias de los programadores, entonces los algoritmos aprenden de lo que nosotros les proporcionamos.
- Datos
  - Calidad  
Se necesita una calidad de datos que sea diversa y por ende representativa de la población.
  - De donde provienen  
Se tiene que tomar en cuenta si el proceso de recolección de datos es justo o está sesgado
- Selección de características  
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.
- Grupos de investigación diversos  
Tenemos que considerar ampliar la diversidad de contratación en los campos científicos y programadores.

provocan la discriminación algorítmica con la implementación de algoritmos de machine learning.

## REFERENCIAS

- [1] Angulo, A. M. (2011). "La Pensión de Invalidez en Colombia". *La Pensión de Invalidez en Colombia*. Bogotá D.C., Cundinamarca., Colombia.: Universidad Libre de Colombia.
- [2] Asamblea Nacional Constituyente de Francia. (26 de agosto de 1789). Declaración Universal de los Derechos del Hombre y del Ciudadano.
- [3] Constitución política de Colombia [Const. P.]. (1991). Colombia: Leyer, 2.da ed. 10/02/2020.
- [4] R. Burcio, *Discriminacion Racial a La Mexicana*, CONAPRED, [https://www.conapred.org.mx/index.php?contenido=noticias&id=3766&id\\_opcion=&op=448](https://www.conapred.org.mx/index.php?contenido=noticias&id=3766&id_opcion=&op=448) (Accedido el 26 de Agosto de 2022).
- [5] *Una de cada cinco personas de 18 años y más, declaró haber sido discriminada, Encuesta Nacional sobre Discriminacion (ENADIS), 2017*

## IV. CONCLUSIÓN

Se tiene que poner más atención en cómo se recolectan, procesan y organizan los datos, para que estos no lleven 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 que 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