



Sesgos en los algoritmos basados en "machine learning": caso de estudio sobre clasificación de imágenes profesionales

Autores:

Javier Blanco Sánchez
Adrián Mestas Bustamante
Aitor Pico García
Mateo Rama García

Tutor:

Daniel Sánchez Repullo

Código:

C201L423



Autores: Blanco Sánchez, Javier; Mestas Bustamante, Adrián; Pico García, Aitor; Rama García, Mateo.

Resumen

El continuo desarrollo de las nuevas tecnologías conlleva un uso cada vez más extendido de la inteligencia artificial. Se recurre a ella para resolver problemas repetitivos o tediosos. Esto ha permitido mejorar los motores de búsqueda, realizar reconocimiento facial, desarrollar asistentes virtuales... El peligro surge cuando se delega la toma de decisiones importantes, como podría ser un proceso de selección de personal, sin entender realmente cómo funciona. Es necesario que dichas decisiones estén libres de sesgos o, al menos, conocer cuáles son los casos donde la inteligencia artificial pueda ser discriminatoria. Nuestra propuesta trata de replicar el proceso a pequeña escala, poniendo de relevancia la importancia de detectar, analizar las causas y tratar de evitar los sesgos de la inteligencia artificial.

Palabras clave: género, inteligencia artificial, IA, conductas, racismo, sesgos, discriminación

Overview

The continuous development of new technologies leads to increasingly widespread use of artificial intelligence (AI), which is used to solve repetitive and tedious problems and it has resulted in better search engines, facial recognition, virtual assistants, and more. Danger appears when important decision-making is delegated, such as a personnel selection process, without really understanding how it works. It is necessary that these decisions are free of bias or, at least, know which are the cases where AI can be discriminatory. Our proposal tries to replicate the process on a small scale, emphasizing the importance of detecting, analyzing the causes, and trying to avoid the biases of AI.

Keywords: gender, artificial intelligence, AI, conducts, racism, biases, discrimination.

ÍNDICE

1.	Introducción	3
2.	Objetivos	4
2.1.	Objetivo general	4
2.2.	Objetivos específicos	5
3.	Procedimiento experimental	6
3.1.	Planteamiento del problema	6
3.2.	Búsqueda de imágenes	7
3.3.	Entrenamiento de la Inteligencia Artificial	8
3.4.	Casos de prueba y recopilación de datos	9
3.5.	Análisis de datos	10
3.6.	Exposición de los datos y conclusiones	11
4.	Resultados y discusión	11
5.	Conclusiones	24
6.	Limitantes de la investigación	24
7.	Futuras líneas de investigación	25
8.	Bibliografía	26

1. Introducción

En estos tiempos donde la gran mayoría del mundo tiene acceso a un móvil con conexión a internet, el mercado digital es uno de los mejor valorados y a la vez más amplios del mundo. Un hábitat natural para que campos como la inteligencia artificial se desarrollen, puesto que nos permite realizar infinidad de tareas, desde las más monótonas y sencillas, a las más complejas; en tiempos extremadamente pequeños.

Pero no todo es positivo. Una mala implementación en el diseño del algoritmo que subyace como base de toda inteligencia artificial puede acarrear graves consecuencias. Dependiendo del problema que se desee resolver, un fallo de programación podría pasar inadvertido. De hecho, es lo más habitual, pues las personas que programan la inteligencia no son las que se encargan de aplicarla a la resolución de un determinado problema.

El proceso de prueba es un proceso de caja negra donde no interesa el código, interesa entender si a una determinada entrada le corresponde una determinada solución. Al menos es así en los casos más simples. A medida que el problema es más complejo, entender si la solución es adecuada o no, no resulta tan sencillo. Si el propósito asignado a la inteligencia artificial es, por ejemplo, el de filtrar personas para un posible puesto de trabajo no es tan claro quién es el candidato óptimo. Es aquí cuando la estadística, como disciplina científica, se revela esencial. Gracias a ella, es posible conocer si la inteligencia artificial entrenada dispone o no de algún tipo de sesgo.

En nuestra investigación nos hemos encargado de entrenar una inteligencia artificial con los primeros resultados que aparecían en Google al realizar una búsqueda de una profesión. De tal forma que, cuando aparecían pocos resultados de algún género, se entrenaba el modelo con una pobre representación de este. No obstante, si no aparecía ninguna foto de un determinado género, sí que intentábamos añadir alguna con el objeto de que los resultados fueran más realistas.

Básicamente, nuestra intención fue simular las condiciones de una inteligencia artificial promedio, es decir, entrenada con fotos de mujeres y hombres no siempre en la misma proporción ni cantidad. Con esta inteligencia artificial así entrenada,

nuestra labor era analizar la existencia de sesgos y si los motivos de esos posibles sesgos eran debidos a la selección de casos con los que se entrenó al modelo.

2. Objetivos

Los objetivos planteados a la hora de realizar el presente trabajo poseen motivaciones en dos planos diferenciados. El objetivo general es de índole exógeno, pues viene impuesto por nuestro profesor de matemáticas, quien ha tutorizado nuestra labor. Mientras que los objetivos específicos son de orden endógeno, pues fueron establecidos por nosotros.

2.1. Objetivo general

El objetivo general no es otro que alcanzar algunos de los contenidos para la asignatura de Matemáticas I recogidos en el Decreto 42/2015, de 10 de junio, por el que se regula la ordenación y se establece el currículo del Bachillerato en el Principado de Asturias. En concreto, se pretende por nuestra parte:

- Elaborar un informe científico escrito que recoja el proceso de investigación realizado, con el rigor y la precisión adecuados.
- Planificar adecuadamente el proceso de investigación, teniendo en cuenta el contexto en que se desarrolla y el problema de investigación planteado.
- Desarrollar procesos de matematización en contextos de la realidad cotidiana (numéricos, geométricos, funcionales, estadísticos o probabilísticos) a partir de la identificación de problemas en situaciones de la realidad.
- Desarrollar y cultivar las actitudes personales inherentes al quehacer matemático.
- Superar bloqueos e inseguridades ante la resolución de situaciones desconocidas.
- Reflexionar sobre las decisiones tomadas, valorando su eficacia y aprendiendo de ellas para situaciones similares futuras.
- Emplear las herramientas tecnológicas adecuadas, de forma autónoma, realizando cálculos numéricos, algebraicos o estadísticos, haciendo representaciones gráficas, recreando situaciones matemáticas mediante

simulaciones o analizando con sentido crítico situaciones diversas que ayuden a la comprensión de conceptos matemáticos o a la resolución de problemas.

- Utilizar las Tecnologías de la Información y la Comunicación de modo habitual en el proceso de aprendizaje, buscando, analizando y seleccionando información relevante en internet o en otras fuentes, elaborando documentos propios, haciendo exposiciones y argumentaciones de estos y compartiendo estos en entornos apropiados para facilitar la interacción.
- Describir y comparar conjuntos de datos de distribuciones bidimensionales, con variables discretas o continuas, procedentes de contextos relacionados con el mundo científico y obtener los parámetros estadísticos más usuales, mediante los medios más adecuados (lápiz y papel, calculadora, hoja de cálculo) y valorando la dependencia entre las variables.
- Interpretar la posible relación entre dos variables y cuantificar la relación lineal entre ellas mediante el coeficiente de correlación, valorando la pertinencia de ajustar una recta de regresión y, en su caso, la conveniencia de realizar predicciones, evaluando la fiabilidad de estas en un contexto de resolución de problemas relacionados con fenómenos científicos.
- Utilizar el vocabulario adecuado para la descripción de situaciones relacionadas con la estadística, analizando un conjunto de datos o interpretando de forma crítica informaciones estadísticas presentes en los medios de comunicación, la publicidad y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.

2.2. Objetivos específicos

Los objetivos secundarios son los siguientes:

- Identificar los potenciales sesgos que se producen en toda investigación científica.
- Evaluar el funcionamiento de una inteligencia artificial entrenada para la clasificación de imágenes mediante algoritmos de “machine learning”.
- Establecer factores que afecten a la eficiencia de una inteligencia artificial.

3. Procedimiento experimental

La propuesta del proyecto comenzó cuando nuestro profesor de matemáticas, Daniel, nos propuso la posibilidad de desarrollar un informe estadístico basado en los sesgos de la inteligencia artificial, con el fin de poder analizar cómo varían las decisiones de las inteligencias artificiales desarrolladas por los humanos.

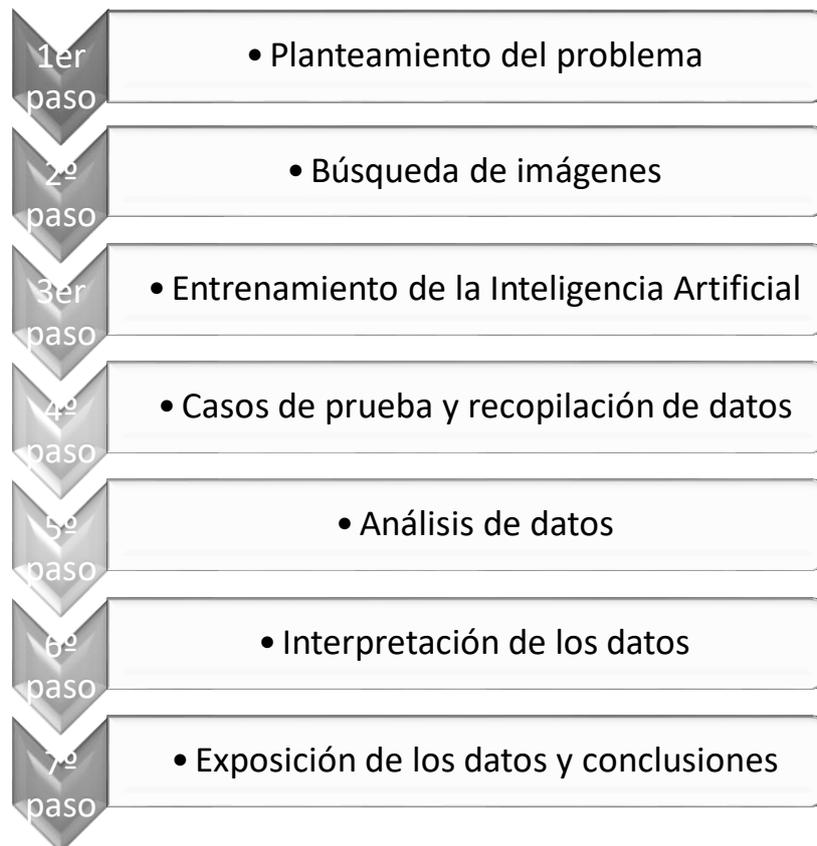


Ilustración 1. Esquema del procedimiento experimental (Elaboración propia)

Los autores del presente trabajo nos comprometimos con el desarrollo de esta idea, ya que consideramos que este trabajo puede reflejar y mejorar los factores principales que decidirán en los próximos años las oportunidades dictadas por las decisiones artificiales.

3.1. Planteamiento del problema

La primera dificultad que encontramos fue el planteamiento del problema. Nuestro profesor nos comentó las ventajas y desventajas de la tecnología basada en el “machine learning” y hasta qué punto nos afecta en nuestra vida diaria. Más tarde nos mostró una herramienta web en la que podíamos entrenar nuestra propia

inteligencia artificial. Tras un breve ejemplo sobre el reconocimiento en tiempo real de si un individuo portaba o no mascarilla, nos decidimos a realizar un informe recopilando, detallando y documentando los sesgos de la inteligencia artificial y su respuesta ante la elección de profesiones en base a un juego de fotos previamente seleccionado.

Es importante saber que esta tarea encomendada no nos ha supuesto solamente un trabajo académico sino una forma de aprender cómo está funcionando y cómo funcionará nuestro mundo en los próximos años; especialmente el apartado laboral,



Foto 1. Presentación del proyecto a la clase.

que es en el que más hemos indagado y que consideramos el comienzo de la puesta en práctica de la inteligencia artificial en las decisiones sociales.

3.2. Búsqueda de imágenes

Para realizar una búsqueda de imágenes efectiva, planteamos la posibilidad de dividirnos las profesiones y sus correspondientes imágenes. Cada uno de nosotros buscamos por internet las imágenes de 6 profesiones distintas con las que poder

comenzar a entrenar la inteligencia artificial. Concluimos que lo más oportuno, por cuestiones de optimización de tiempo, era seleccionar 30 imágenes diferentes por cada trabajo intentado variar lo máximo posible la vestimenta, la posición de los sujetos en las imágenes y las diferencias personales que presentaba cada una de ellas. De esta manera, la inteligencia artificial podría reconocer un número de sujetos de prueba suficientemente alto, presentando menos problemas a la hora de asignar los porcentajes correspondientes.

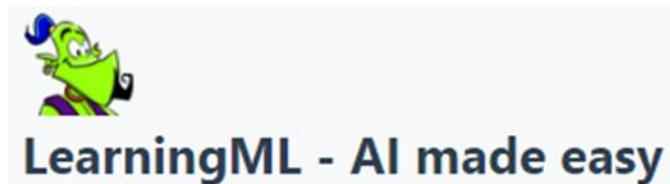


Foto 2. Logo de Learning ML

3.3. Entrenamiento de la Inteligencia Artificial

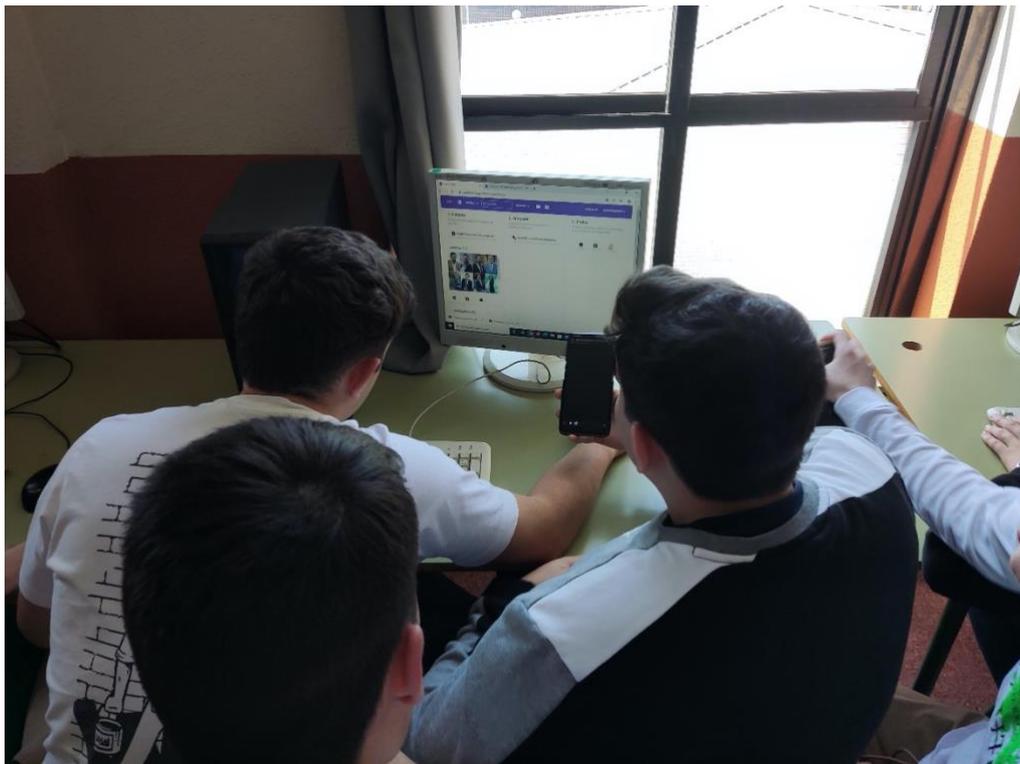


Foto 3. Entrenando la inteligencia artificial en clase

Con todo lo anterior preparado era momento de entrenar la inteligencia artificial. Tras una exhaustiva búsqueda, en internet y otras fuentes, de distintos tipos de inteligencias artificiales, decidimos que la más adecuada para nuestro trabajo era

learningML ya que su proceso de aprendizaje resultaba intuitivo y asequible para todos los niveles, su procesamiento de datos era rápido y se podían utilizar imágenes, lo cual era fundamental y nos facilitó el entrenamiento de la inteligencia artificial. Gracias a esto, la tarea se reducía únicamente a importar las más de 600 imágenes; afortunadamente, era tan fácil como crear el empleo y subirlas todas de una sola vez.

3.4. Casos de prueba y recopilación de datos

Cuando tuvimos entrenada y preparada la inteligencia artificial para poder empezar con los casos de prueba, nos volvimos a repartir las profesiones y buscamos otras 6 fotos por cada profesión, con el objetivo de poder experimentar con distintas personas y objetos para poder ver los resultados de cada foto. Cada miembro del equipo se encargó de sus profesiones y de sus fotos, que más tarde fueran puestas en común junto a los porcentajes correspondientes de cada profesión.

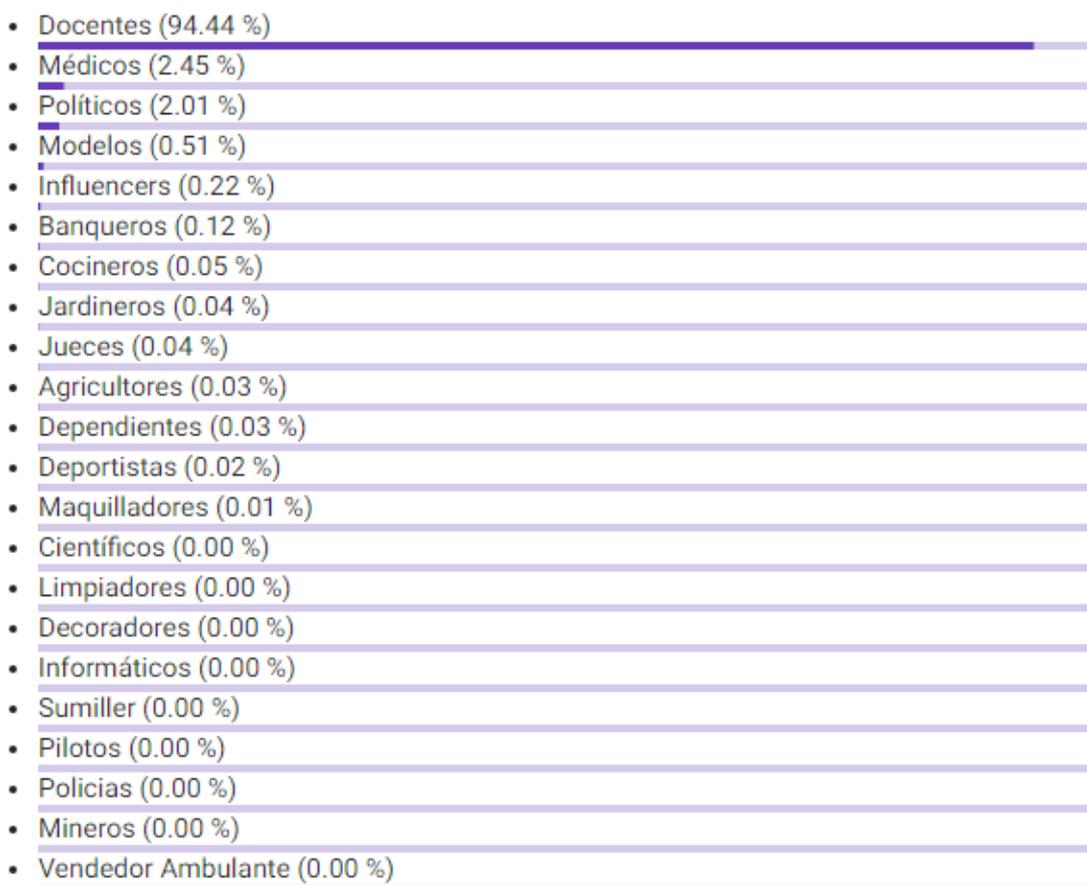


Foto 4. Porcentajes de asignación a cada profesión proporcionados por la inteligencia artificial al someter al modelo una determinada foto de prueba

Las fotos consistieron en un hombre y mujer blanca, un anciano, una persona perteneciente a otra etnia, un fondo acorde a la profesión o lo más descriptivo posible y, por último, un objeto de dicha profesión.

Para poder realizar los posteriores gráficos, fue clave la organización del equipo. Primero necesitábamos recopilar los datos en un documento de Excel. Estos datos eran devueltos por la inteligencia artificial en forma de un porcentaje que representaba la posibilidad de pertenecer a cada una de las profesiones (Foto 4). Fue necesario recopilar un gran número de datos, puesto que por cada profesión había 6 casos de prueba y por cada caso de prueba 29 porcentajes (ya que descartamos una de las profesiones elegidas inicialmente por falta de imágenes suficientes).

Durante el desarrollo del proyecto, la recopilación de los datos era crucial para que no se produjese ningún tipo de alteración a la hora de extraer conclusiones o analizar los datos por lo que estuvimos comunicándonos constantemente para que no hubiese ningún error. A parte se debe destacar que, después de haber recopilado todos los datos, una persona del equipo se encargó de revisar que los datos estuviesen en sus casillas correspondientes y en el formato adecuado.

3.5. Análisis de datos

El análisis de datos se puede dividir en dos partes: los porcentajes de acierto y tendencia de la inteligencia artificial en general y por profesión (foto 5), y las gráficas por caso de prueba de dicha profesión.

Durante este proceso, para una mayor rapidez y un mejor análisis de los datos, nos ayudamos de los gráficos que nos proporciona la herramienta de Excel. Para comprobar la existencia de posibles sesgos, representamos conjuntamente los porcentajes asignados a cada uno de los rasgos distintivos del modelo de prueba para cada profesión; así como la distribución de los porcentajes de acierto por cada factor o diagramas de dispersión al tratar de relacionar dos variables diferentes.

Nos repartimos nuevamente el trabajo entre los 4 miembros del equipo buscando una distribución uniforme y equitativa para una mayor eficiencia a la hora de interpretar los datos y así poder extraer unas conclusiones más acertadas.

	Rasgo Étnico Informático	Fondo Predominante Informático	Elemento Predominante Informático
Informáticos	18.82	99.11	75.58
Deportistas	0.02	0	0.02
Agricultores	0	0	0.03
Banqueros	0.41	0.01	0.02
Científicos	0	0	0.01
Cocineros	0	0.01	0.31
Decoradores	0.18	0.82	3.14
Dependientes	57.74	0.01	11.59
Docentes	0	0	1.98
Influencers	13.72	0.03	0.03
Jardineros	0	0	6.29
Jueces	0.15	0	0.03
Limpiadores	0	0	0.02
Maquilladores	0.76	0	0
Médicos	0	0	0.01
Mineros	0	0	0.01
Modelos	0.03	0	0
Pilotos	0.01	0.01	0.04
Policías	0	0	0
Políticos	0	0	0
Sumiller	8.16	0	0
Vendedor Ambulante	0	0	0.119
	100	100	99.229

Foto 5. Representación mediante una tabla de Excel del porcentaje de cada profesión asignado por la inteligencia artificial a tres modelos de prueba.

3.6. Exposición de los datos y conclusiones

Una vez que tuvimos todos los datos analizados en gráficos, nos dispusimos a plasmar nuestras conclusiones en el presente informe.

4. Resultados y discusión

A continuación, expondremos los datos recabados en función de cómo clasificaba la inteligencia artificial las fotos seleccionadas en relación con distintos parámetros para una profesión, como el sexo, la edad, los rasgos étnicos, un fondo/escenario predominante y objeto distintivo.

Para la mayoría de los casos la inteligencia fue capaz de clasificar adecuadamente la profesión de las personas que aparecían en las fotos de prueba con elevado porcentaje de exactitud. Pero como era de esperar, hay bastantes excepciones en las que el algoritmo no ha sido capaz de identificar con precisión la profesión adecuada.

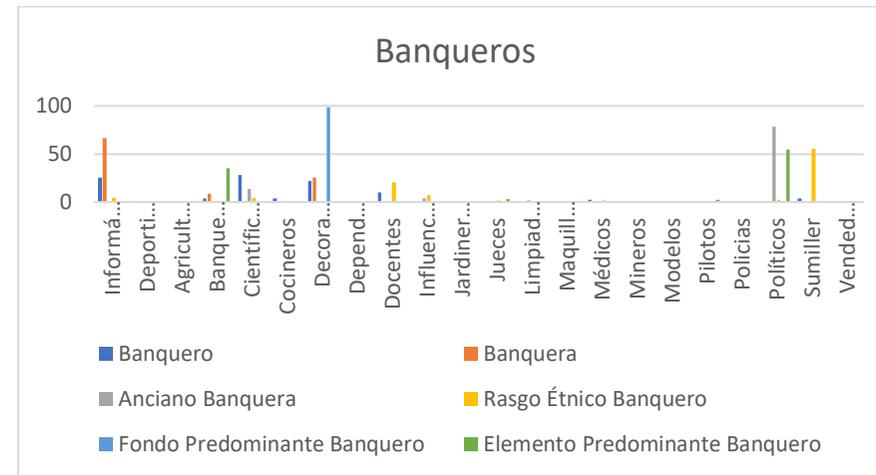
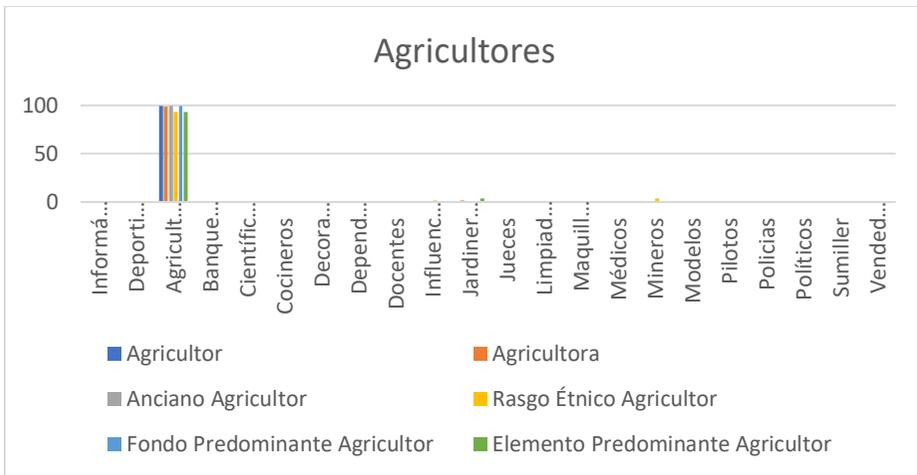
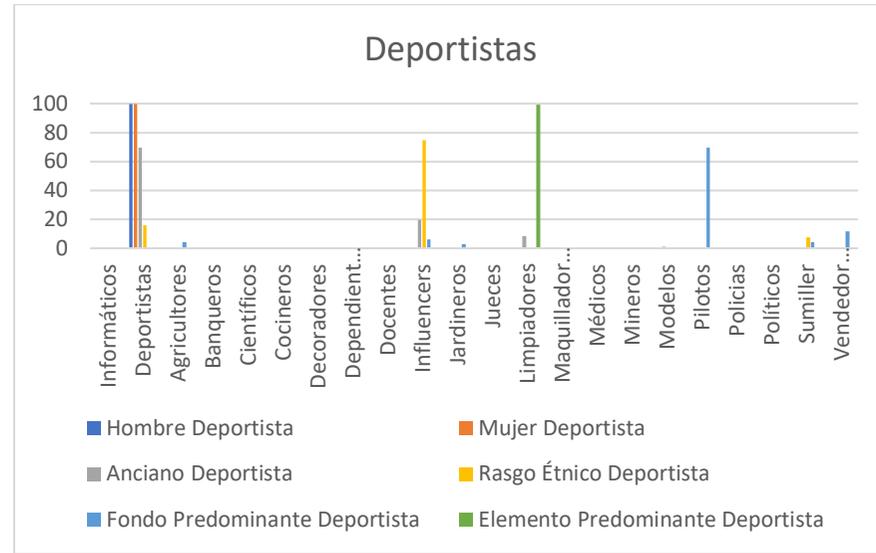
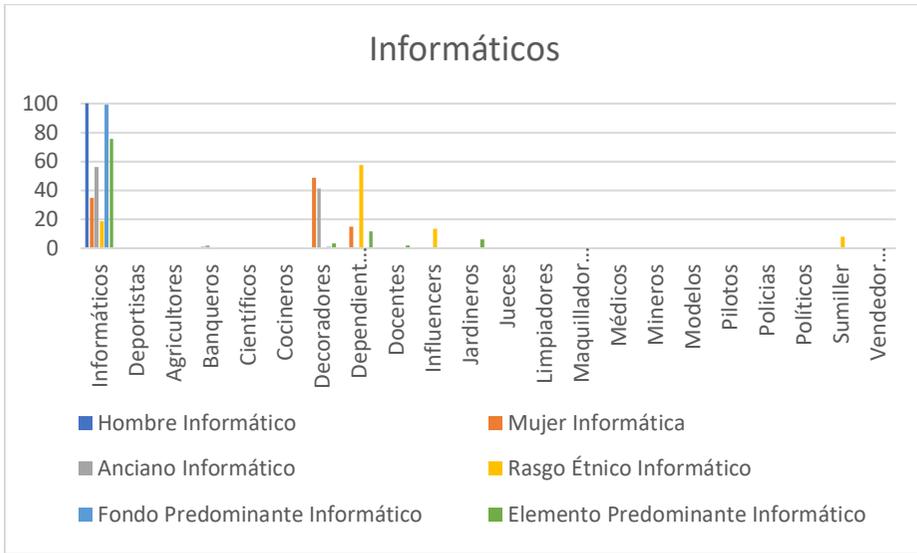
Un caso claro de un buen trabajo realizado por la inteligencia es el de los agricultores (gráfico 3) ya que, para esta profesión, acertó con exactitud la totalidad de las 6 divisiones.

También pueden ser encontrados casos en los cuales no consiguió clasificar correctamente a personas de distintos rasgos étnicos. Prueba de ello son los de los jueces, deportistas, maquilladores, banqueros, modelos, policías, decoradores, dependientes, jardineros, vendedores ambulantes.

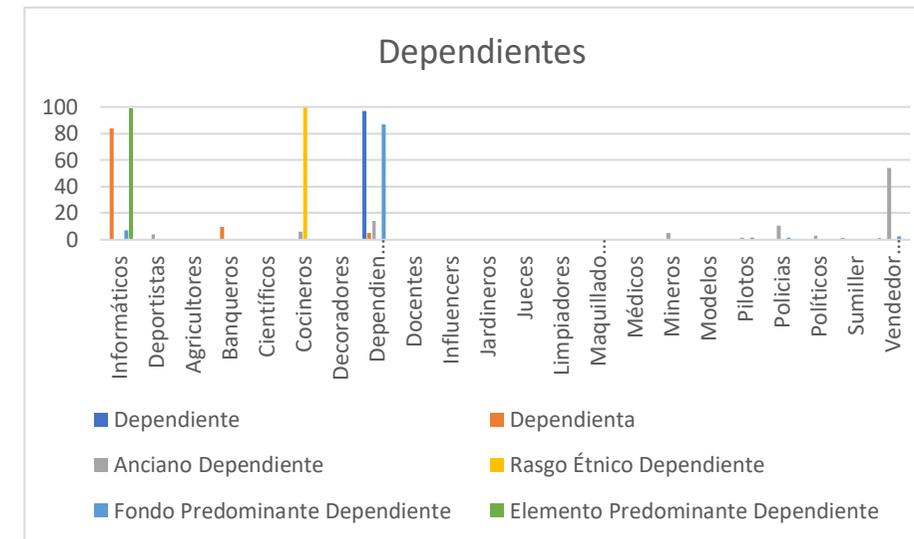
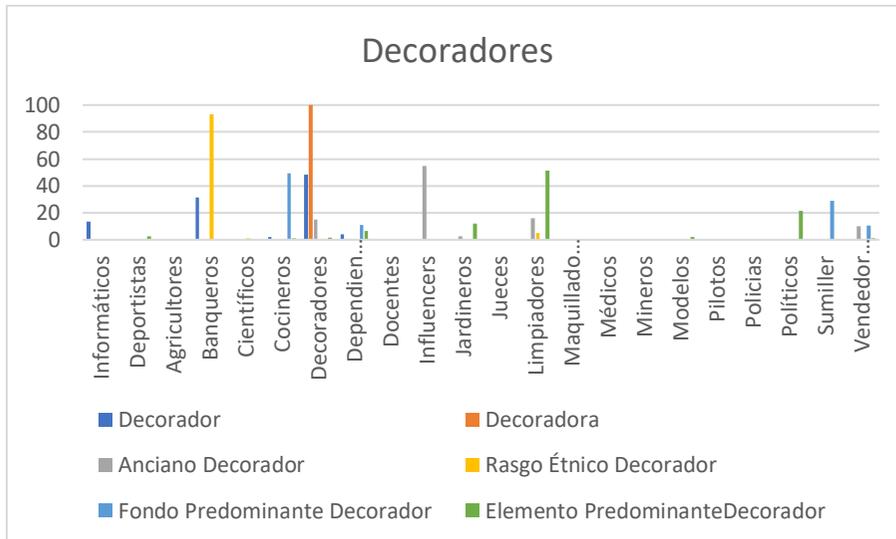
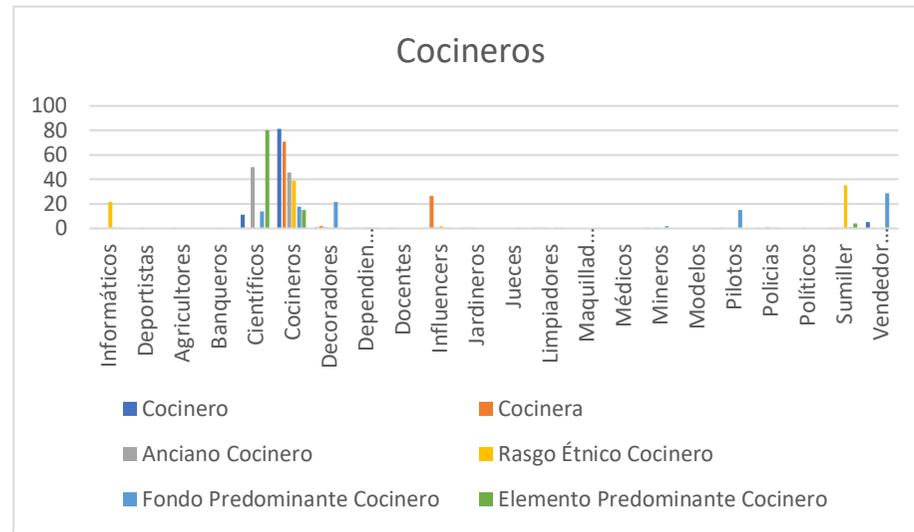
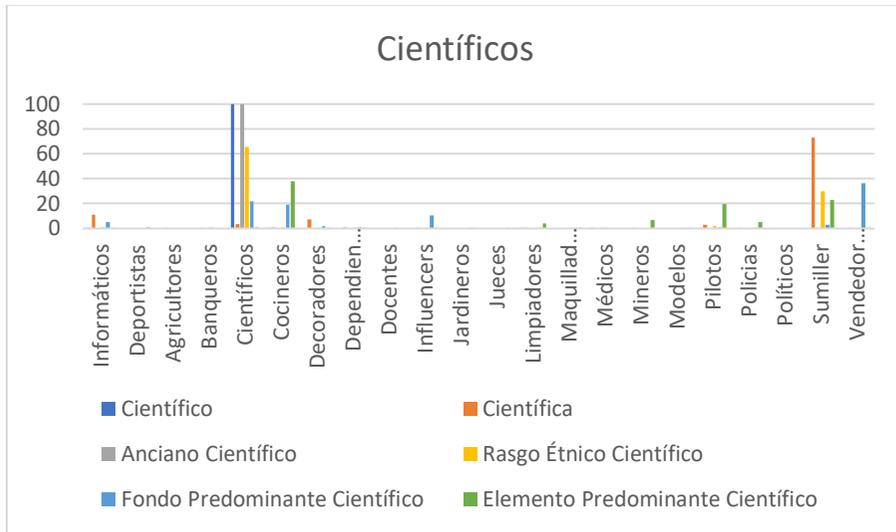
Además, ha habido otros casos en los que ha discriminado a un género: el femenino. Esto ha ocurrido cuando no ha sido capaz de clasificar correctamente a las mujeres en su correspondiente trabajo, como, por ejemplo, banqueros (gráfico 4), dependientes (gráfico 8), influencers (gráfico 10), policías (gráfico 19), ...

El volumen de datos manejados aconseja su presentación mediante gráficos a modo de resumen porque por cada una de las fotos (132) del modelo de prueba se recogieron 22 porcentajes, es decir, un total de 2904 datos.

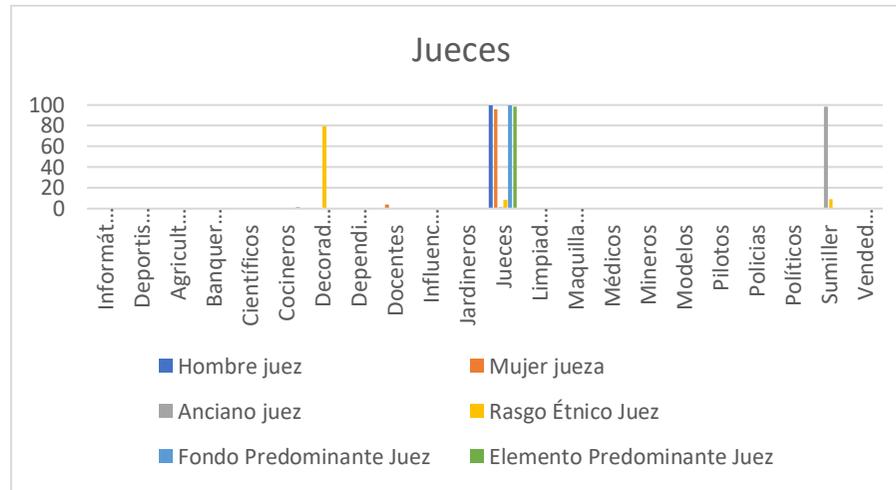
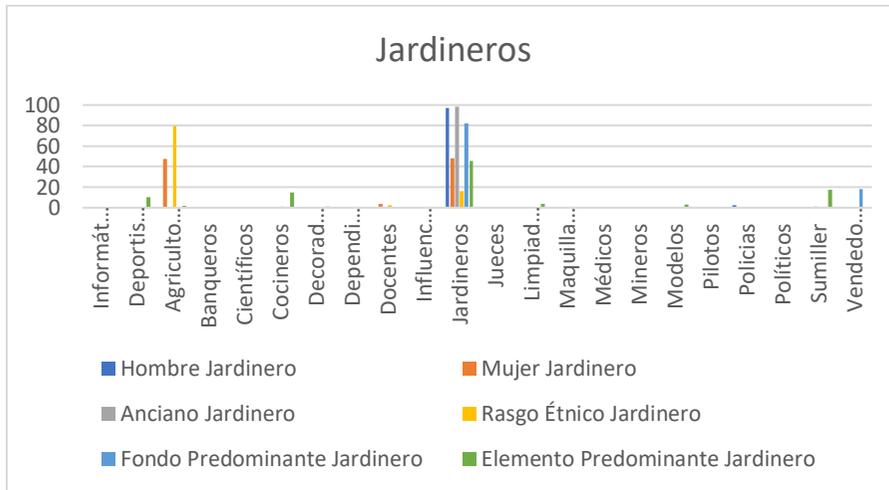
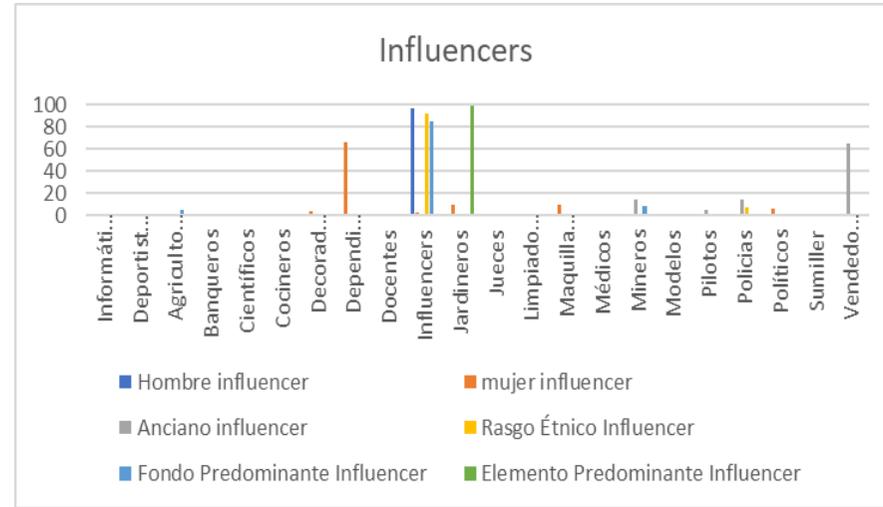
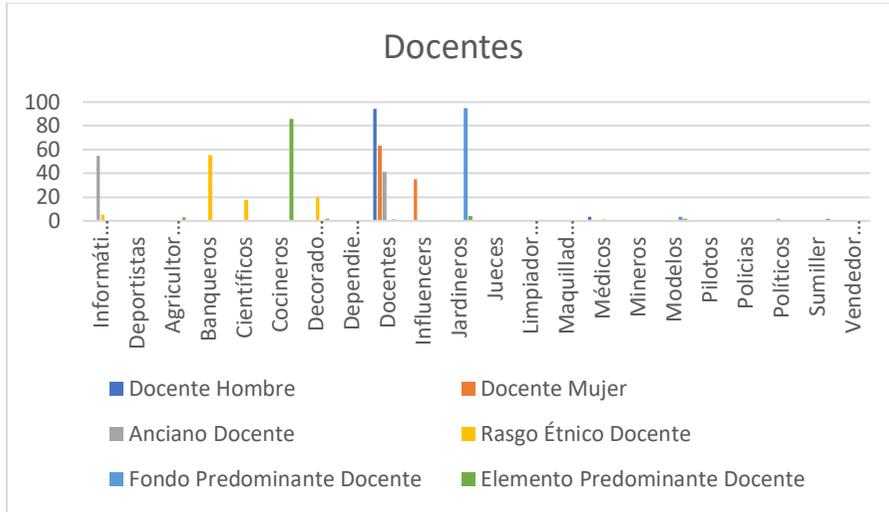
En esta primera serie de gráficos aparece reflejado cómo clasifica el modelo entrenado las fotos de prueba para cada una de las profesiones consideradas.



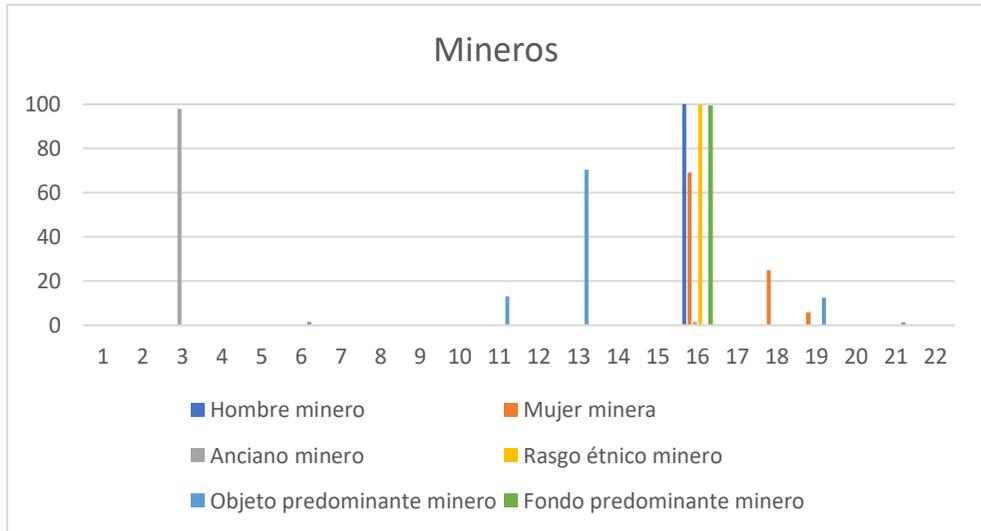
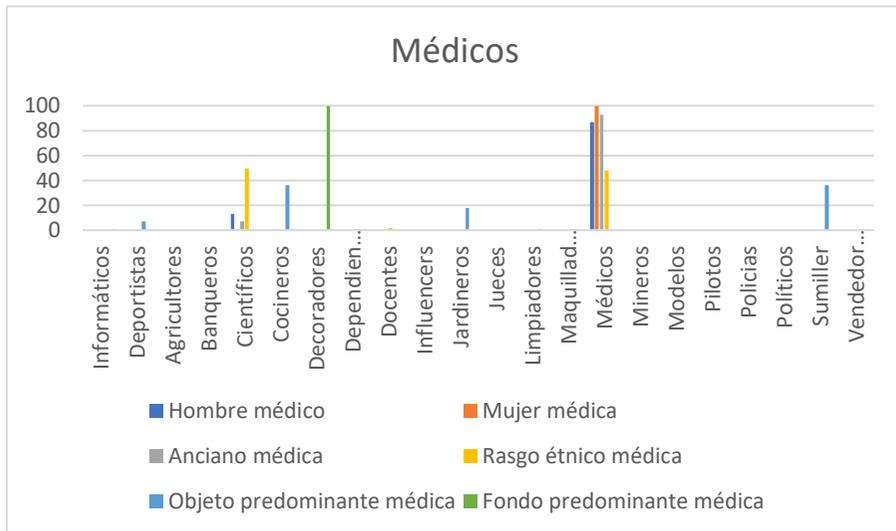
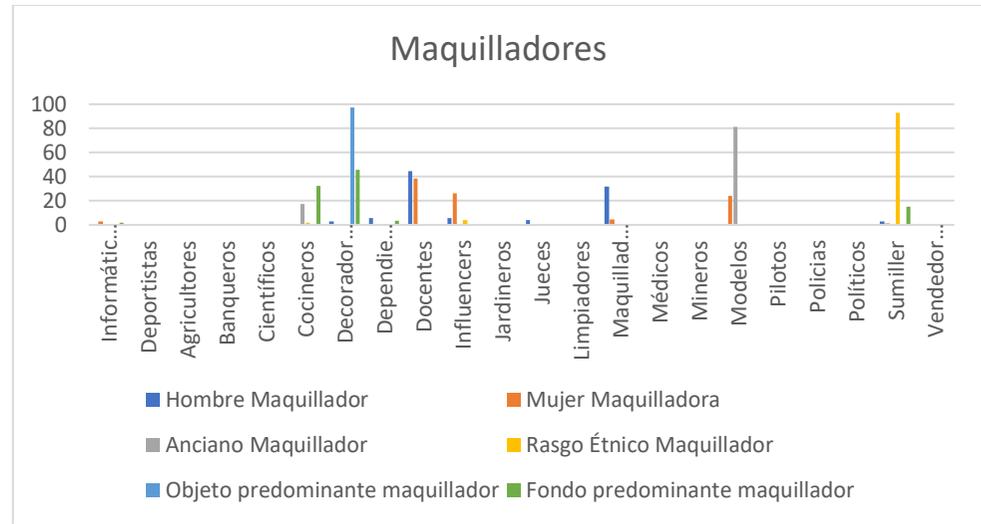
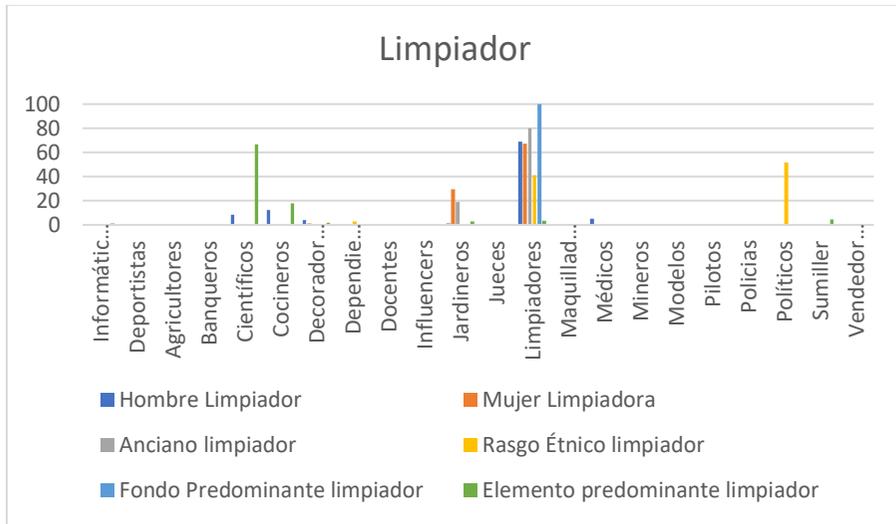
Gráficos 1, 2, 3 y 4. Asignación de profesión de los casos de prueba cuya verdadera profesión viene recogida en el título. (Elaboración propia)



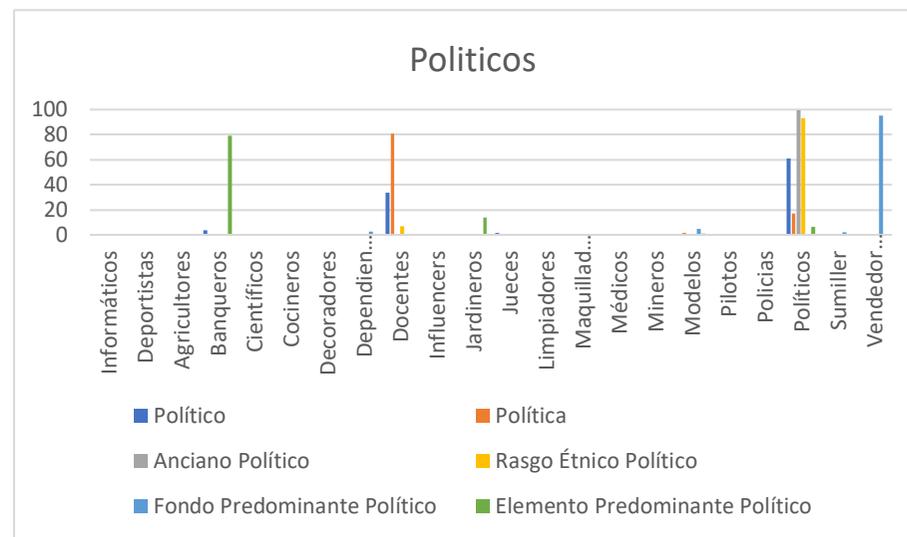
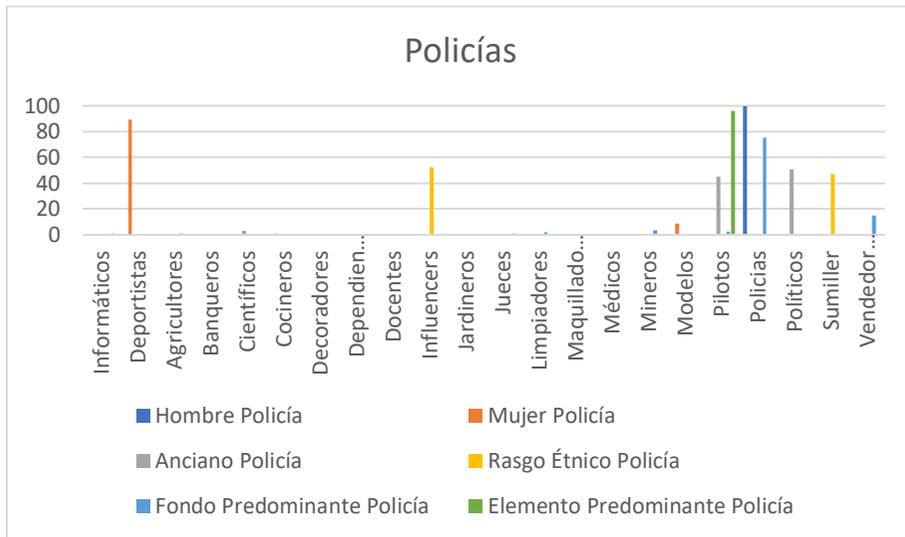
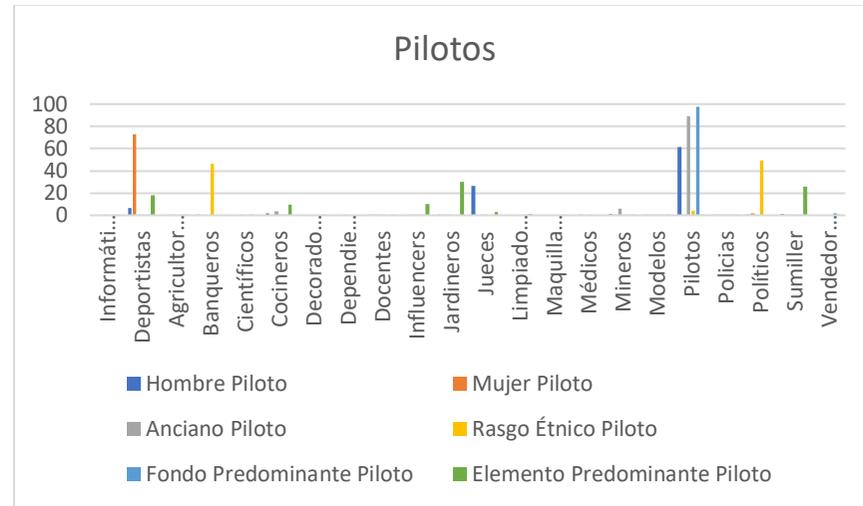
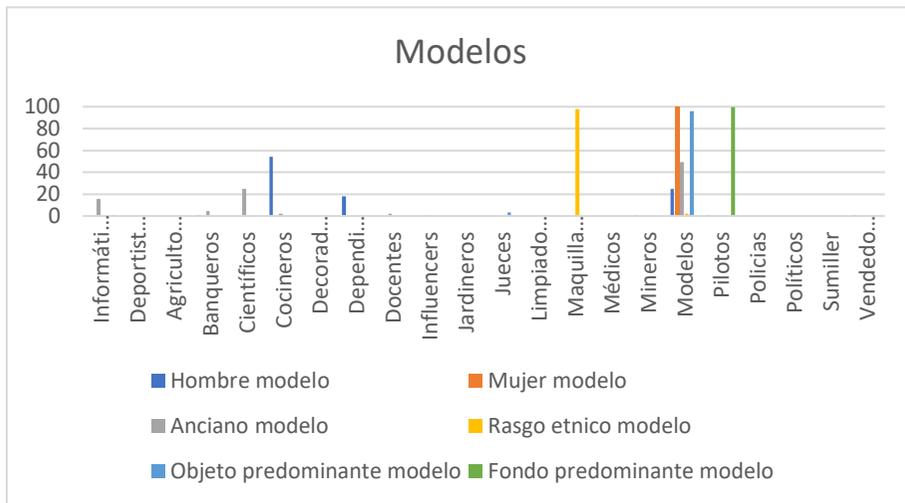
Gráficos 5, 6, 7 y 8. Asignación de profesión de los casos de prueba cuya verdadera profesión viene recogida en el título. (Elaboración propia)



Gráficos 9, 10, 11 y 12. Asignación de profesión de los casos de prueba cuya verdadera profesión viene recogida en el título. (Elaboración propia)



Gráficos 13, 14, 15 y 16. Asignación de profesión de los casos de prueba cuya verdadera profesión viene recogida en el título. (Elaboración propia)



Gráficos 17, 18, 19 y 20. Asignación de profesión de los casos de prueba cuya verdadera profesión viene recogida en el título. (Elaboración propia)

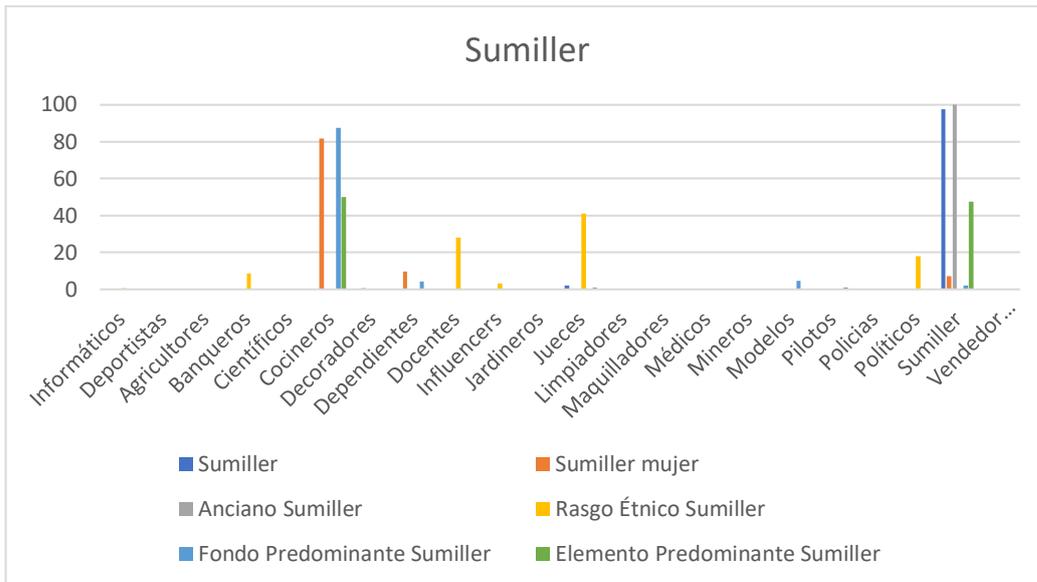


Gráfico 21. Asignación de profesión de los casos de prueba de sumilleres. (Elaboración propia)

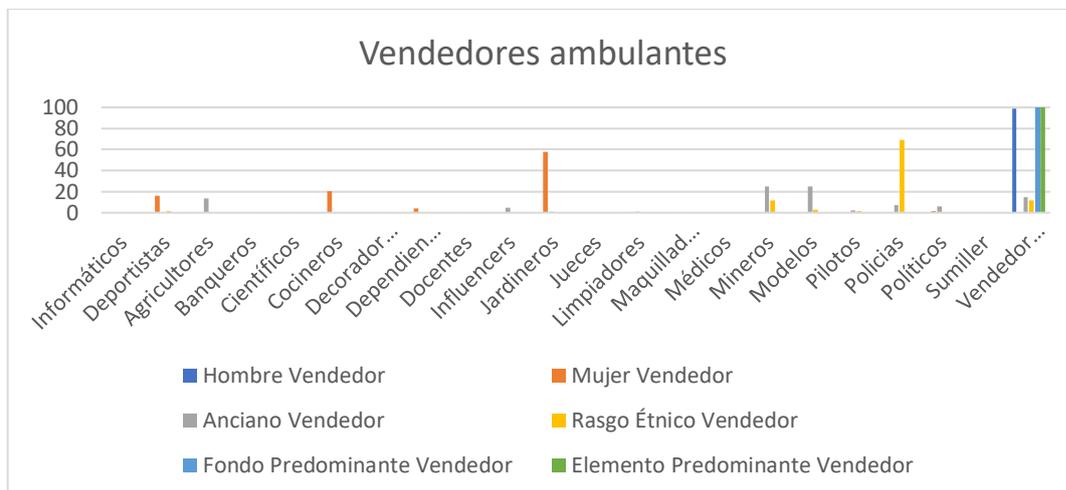


Gráfico 22. Asignación de profesión de los casos de prueba de vendedores ambulantes. (Elaboración propia)

Si analizamos la tasa de éxito de clasificación correcta en función de los factores fijados, destaca sobre manera el caso del sexo masculino frente al resto (Gráfico 23). Tanto es así que hemos decidido centrarnos en el resto del trabajo en intentar explicar las diferencias por razón de género. Pues, como se puede comprobar en la distribución de la tasa de éxito en la correcta clasificación de cada profesión por sexo (Gráfico 24), los diagramas de cajas y bigotes a nivel visual muestran diferencias significativas. La mediana en la caja correspondiente a los varones (la superior) está por encima del 95%, con un recorrido intercuartílico inferior a los 40 puntos; mientras que la mediana en el caso del sexo femenino está en torno al 40% y un recorrido

intercuartílico superior a los 90 puntos. La cruz en el gráfico 24 corresponde a un “outsider”, un elemento extraño que no guarda relación con el resto de la distribución. Ese elemento sería la profesión de policía, pues el porcentaje que le asigna a una foto de un verdadero hombre policía es del 0%.

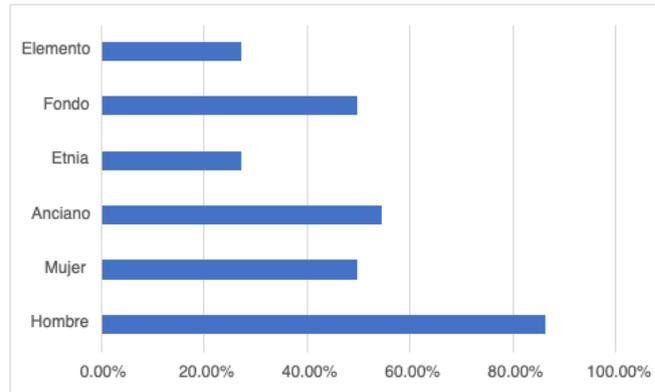


Gráfico 23. Representación de porcentaje de acierto de modelo por factor.

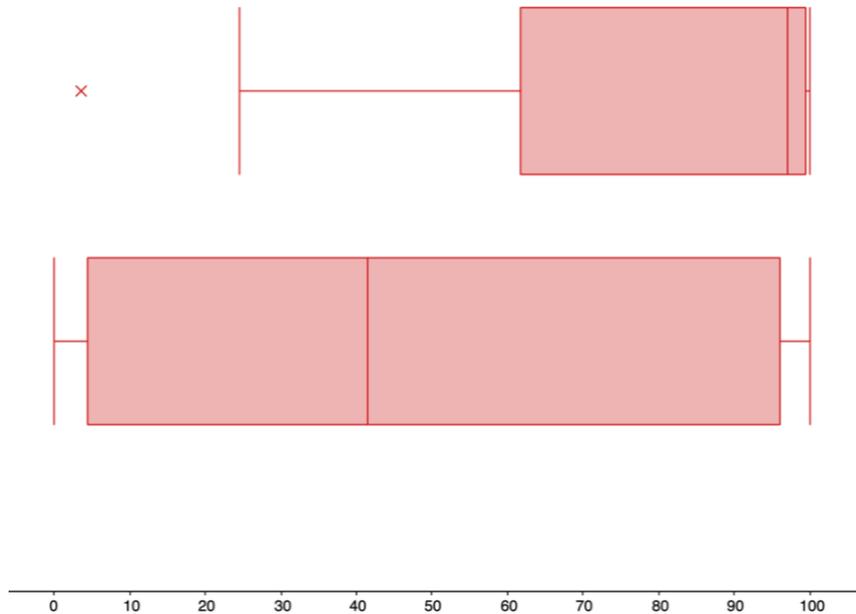


Gráfico 24. Representación de la tasa de éxito por sexo a la hora de elegir profesión.

Caja superior hombres vs caja inferior mujeres.

Para analizar los posibles factores que influyen en esta diferencia, es necesario mostrar los datos extraídos del conjunto de imágenes utilizadas para entrenar el modelo.

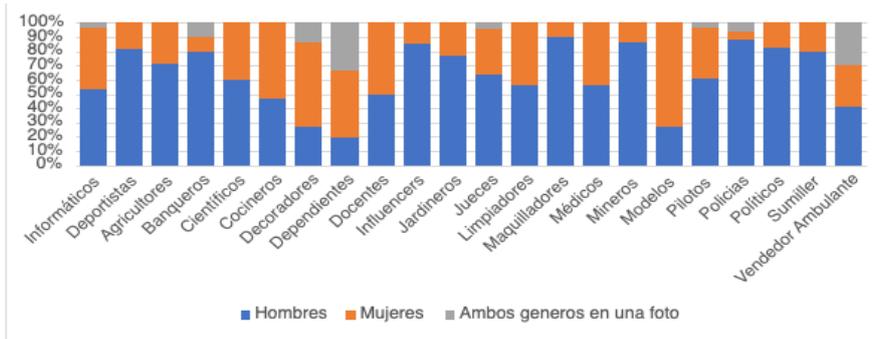


Gráfico 25. Porcentaje de fotos de cada género para cada grupo del entrenamiento.

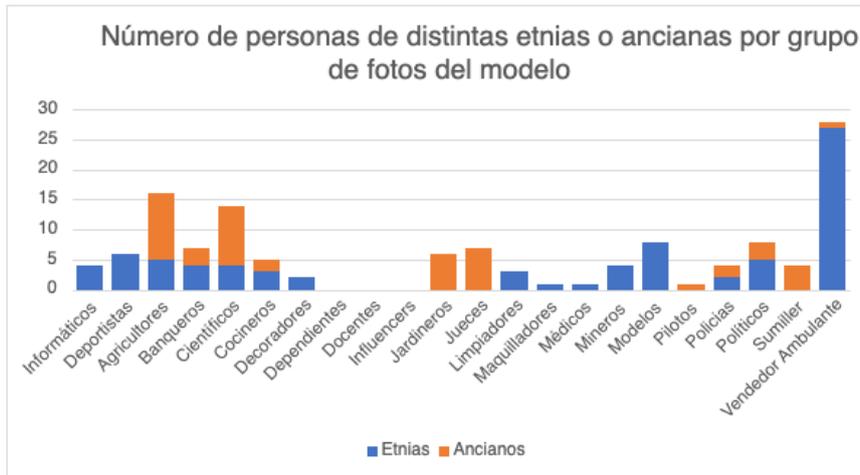


Gráfico 26. Número de fotos con una persona no caucásicas o ancianas por grupo de fotos de modelo.

Partiendo de los datos anteriores, hemos intentado vincular las diferencias en el éxito del modelo a la hora de clasificar las fotos en las que aparecían varones, por un lado, con el porcentaje de fotos de hombres en una profesión sobre el total de hombres (Gráfico 27) y, por otro, con el porcentaje de fotos de hombres en una profesión sobre el número de individuos totales de la población (Gráfico 28). En los gráficos 27 y 28, podemos comprobar, tanto en la nube de puntos como en el valor de R^2 , que no existe ningún tipo de correlación lineal entre esas variables. Análogamente, en los gráficos 29 y 30 se puede observar que tampoco existe una relación de correlación lineal en el caso del género femenino.

En los gráficos 31 y 32 analizaremos la relación existente entre el porcentaje de un tipo de género en una profesión y la probabilidad que asigna el modelo. De los gráficos anteriores y de los coeficientes de correlación allí calculados, se sigue que tampoco podemos establecer ninguna relación entre el número de hombres y mujeres en cada profesión y lo bien que clasifica el modelo fotos de esas profesiones.

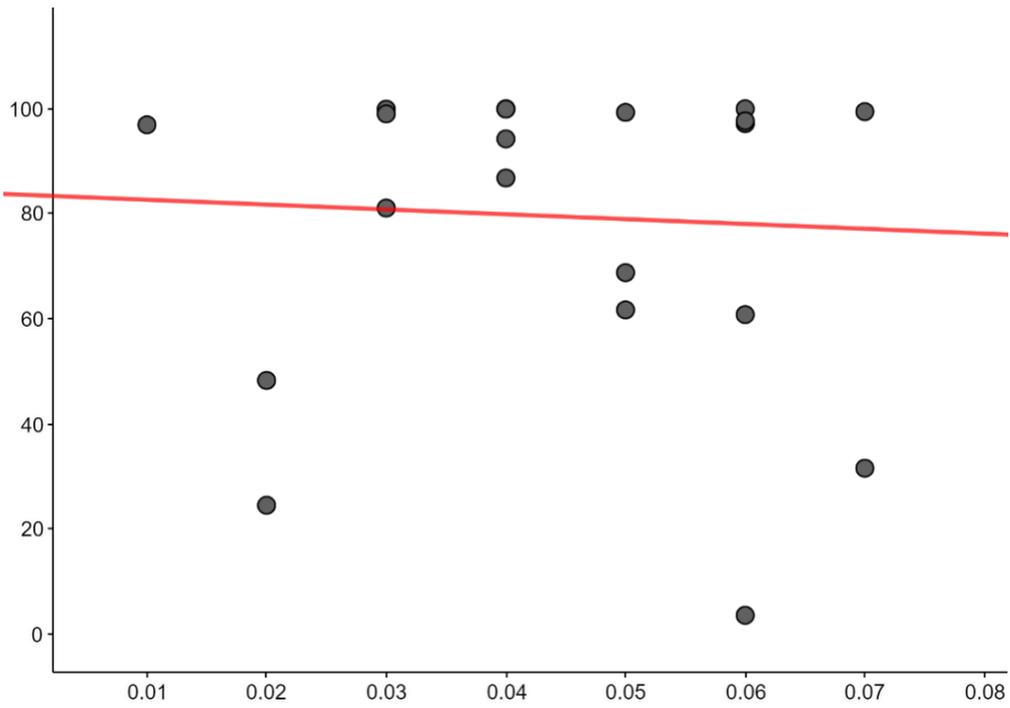


Gráfico 27. Diagrama de dispersión tanto por uno de hombres sobre el número total de ellos en una profesión concreta frente al porcentaje de asignación de modelo. $R^2=0.002$

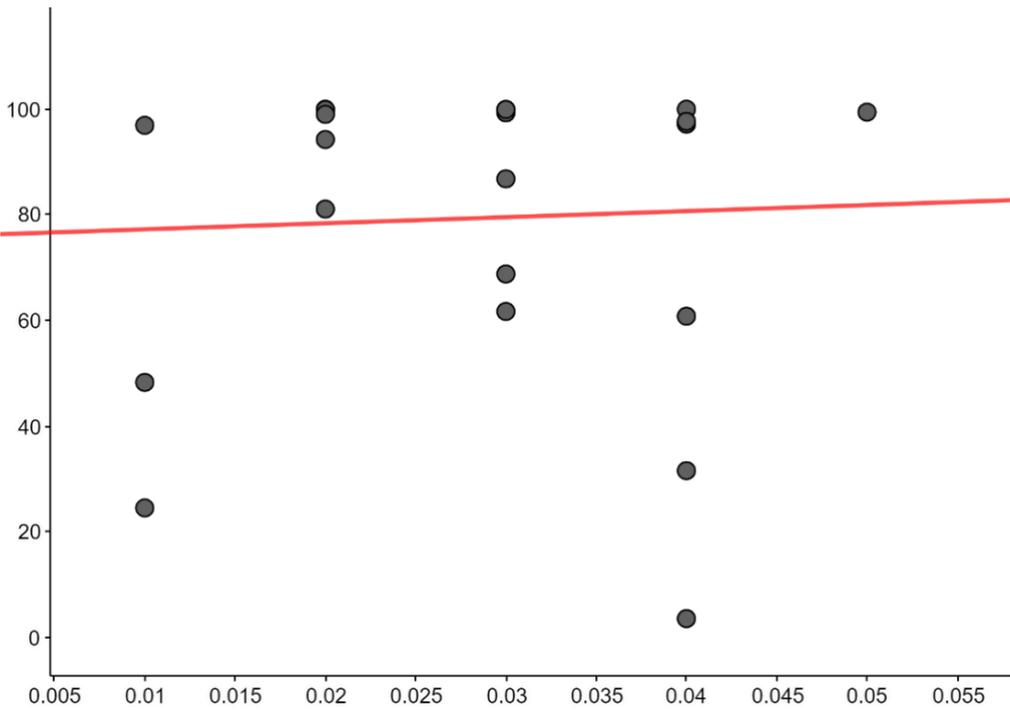


Gráfico 28. Diagrama de dispersión tanto por uno de hombres sobre el número total de individuos frente al porcentaje de asignación de modelo. $R^2=0.0021$

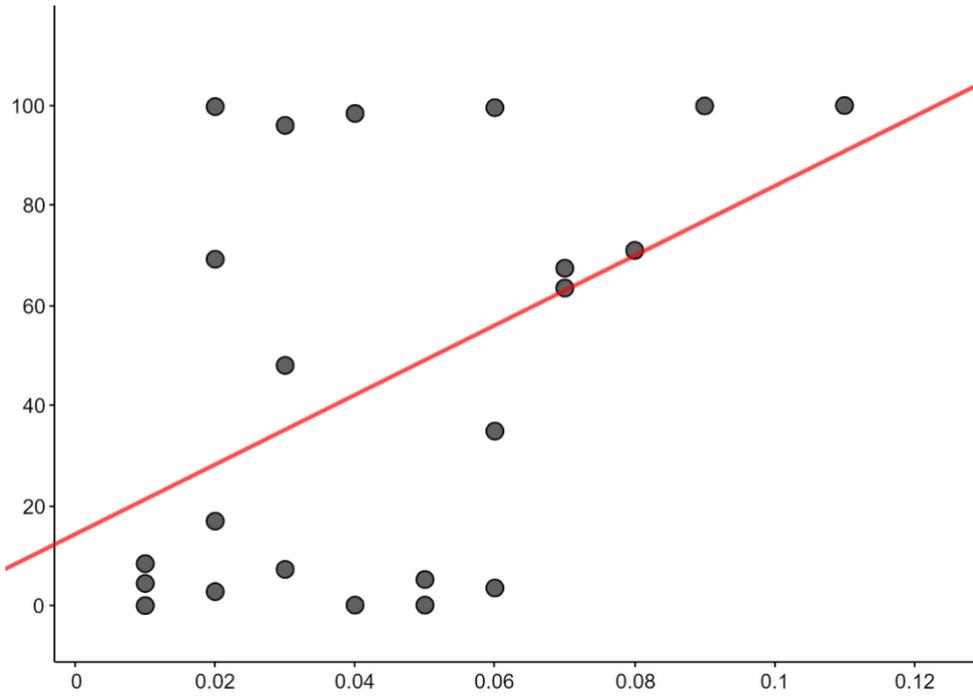


Gráfico 29. Diagrama de dispersión tanto por uno de mujeres en una profesión sobre el total de ellas frente a la probabilidad de asignación de modelo. $R^2=0.2197$

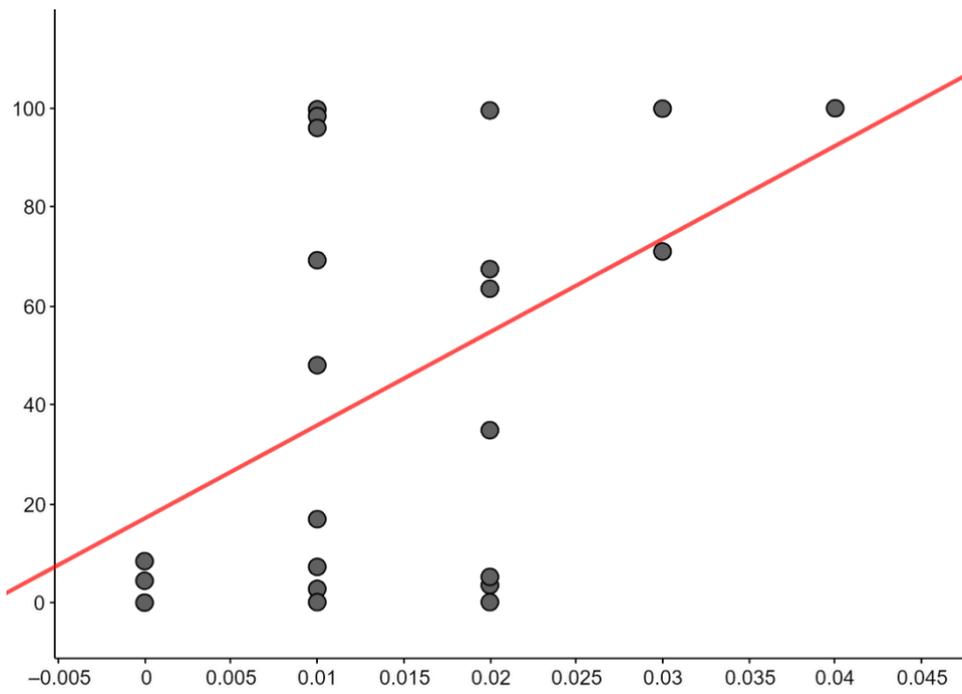


Gráfico 30. Diagrama de dispersión tanto por uno de mujeres sobre el total de personas frente al porcentaje de asignación de modelo. $R^2=0.2117$

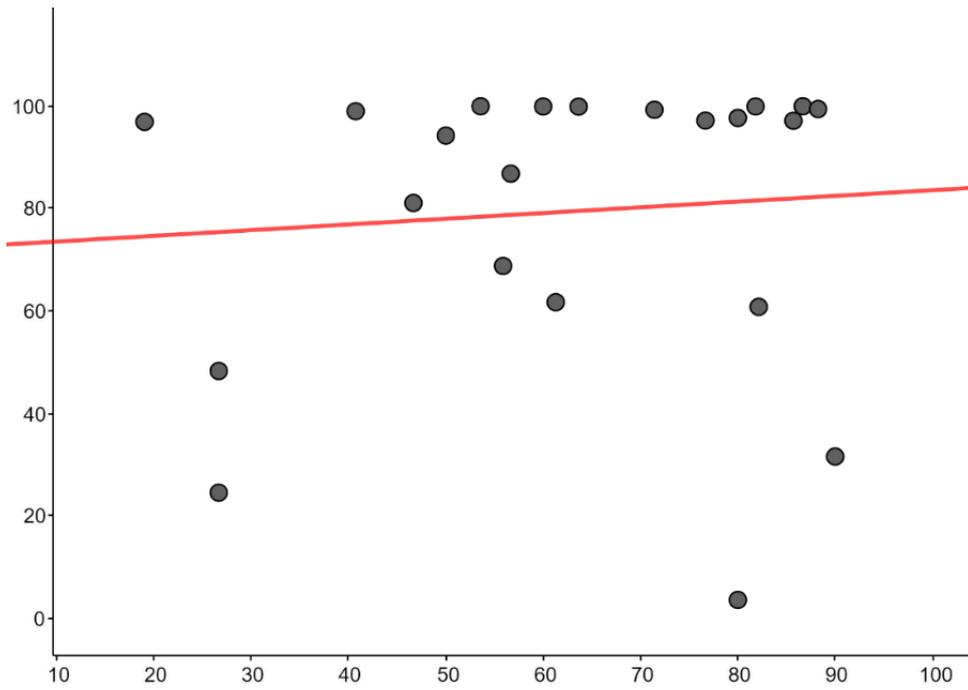


Gráfico 31. Porcentaje de hombres en una profesión frente al porcentaje de asignación de modelo. $R^2=0.0068$

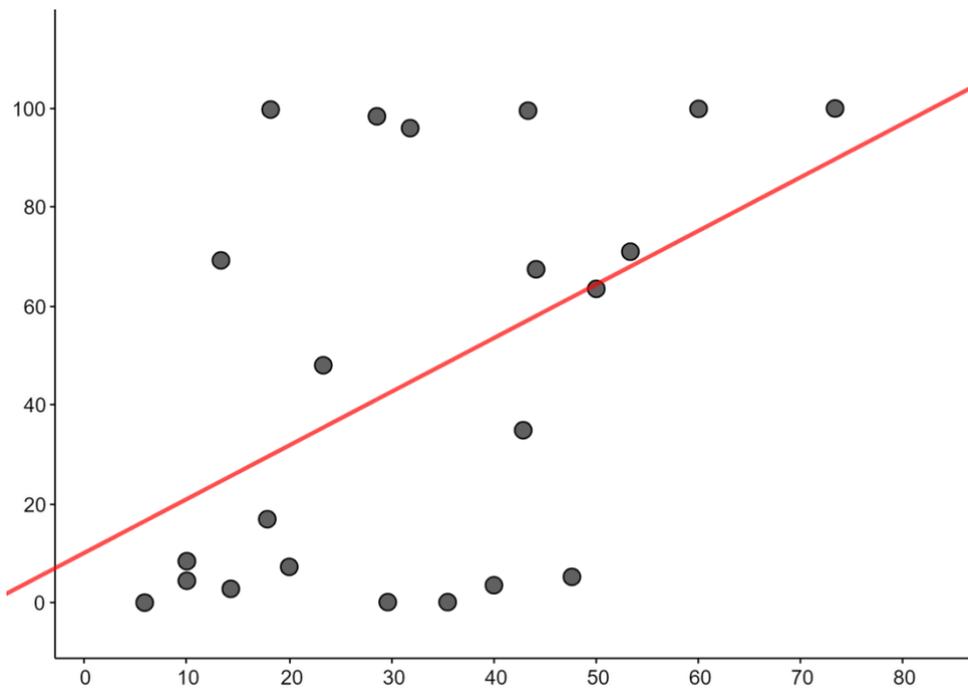


Gráfico 32. Porcentaje de mujeres en una profesión frente al porcentaje de asignación de modelo. $R^2= 0.2265$.

5. Conclusiones

Las principales conclusiones que podemos extraer de la realización del presente trabajo son las siguientes:

- Las inteligencias artificiales requieren de un proceso de entrenamiento que, en principio, podemos controlar. Pero los algoritmos en que se basan hacen que su uso se convierta en un algoritmo de caja negra, es decir, no tenemos ni idea de lo que realmente hacen.
- Podemos estudiar, mediante casos de prueba, los sesgos que se manifiestan en los modelos creados. No obstante, a día de hoy, los conocimientos que poseemos los humildes autores de este trabajo no son suficientes para establecer una correlación estadística entre los elementos que utilizamos para entrenar la inteligencia y los sesgos que provocamos con ellos.
- Así pues, pese a observar que, en prácticamente todos los casos, el modelo que hemos creado clasifica mejor a los hombres que a las mujeres, no sabemos a ciencia cierta el motivo.
- Del mismo modo, podemos especular con los motivos que hacen que ese comportamiento diferencial se produzca. Sin embargo, no hemos podido acercarnos a ellos en el presente estudio.
- No obstante, ni mucho menos podemos afirmar que el trabajo carezca de utilidad alguna. Para empezar porque hemos comprobado en una situación real las limitaciones de nuestros conocimientos estadísticos actuales, lo que nos anima a profundizar en ellos. Para continuar, porque hemos alcanzado todos y cada uno de los objetivos generales que nos habíamos marcado.

6. Limitantes de la investigación

En el apartado de los obstáculos que nos han supuesto un inconveniente a la hora de realizar el informe, son susceptibles de mención los siguientes:

- La irregular selección de fotos de distintas profesiones, en las que debieron ser impuestos unos parámetros iniciales que nos hubieran ayudado a entrenar la inteligencia artificial (tales como un primer plano de cada persona, la elección

acertada de los lugares, contextos y elementos incluidos en las fotos así como la igualdad numérica de personas del mismo sexo, etnia o edad, que sin duda han supuesto alteraciones y condicionantes en el estudio). Ejemplo de ello fue el descarte de la profesión de abogacía debido a la imposible distinción entre las profesiones por parte de la inteligencia artificial.

- Los errores inevitables a la hora de recoger los datos y agruparlos debido al enorme número de cifras manejadas y los fallos técnicos como despistes, equivocaciones o confusiones que surgieron durante el trabajo (significando en ocasiones rehacer partes del informe de nuevo).
- La inexperiencia por nuestra parte a la hora de analizar los datos y resultados obtenidos en la parte estadística, así como la complejidad del reto que nos ha supuesto realizar un estudio que relacione variables estadísticas de forma bidimensional. Lamentablemente, no contamos con los conocimientos necesarios para establecer relaciones multivariantes.
- Nuevamente la inexperiencia en el apartado de redacción del informe, debido a los pocos proyectos similares en los que nos hemos visto envueltos la gran mayoría de miembros que conforman el equipo.
- Finalmente, el tiempo, que ha jugado un papel claro y evidente en nuestra contra, significando días con varias horas de trabajo y falta de preparación previa para el desarrollo del informe.

7. Futuras líneas de investigación

No es ningún secreto que la inteligencia artificial ya forma parte de nuestras vidas, principalmente en las gestiones sociales que se disponen por algoritmos en lugar de personas. Esta tendencia se acentuará en el futuro, ya que cada vez más empresas están invirtiendo más capital en desarrollar nuevas IAs, para ahorrarse costos, simplificar procesos o mejorar el rendimiento.

Este trabajo es solo un pequeño estudio estadístico de las elecciones de una IA sobre las profesiones de diferentes personas en base a un modelo preestablecido, siendo seguramente un sistema que en algún momento será utilizado y entendido como normal. Así pues, el campo de la inteligencia artificial no se limita solo a la idea de la identificación, sino que

existen múltiples campos susceptibles de estudio sobre la implicación en los próximos años de la IA en las decisiones sociopolíticas.

Tras el estudio e investigación llevada a cabo, nos queda constancia de cómo se han desarrollado las distintas fases del trabajo, y cuáles son los métodos utilizados en el mismo, ya sea a la hora de seleccionar datos, entrenar a la inteligencia u otros. En base a las experiencias obtenidas a partir de nuestra investigación, podemos sacar en claro cómo utilizar estos conocimientos a futuro en nuevas investigaciones relacionadas con el ámbito.

Empezando por las diferentes posibilidades que el algoritmo Machine learning ML nos proporcionaba, en su versión beta, cabe destacar que ante futuros proyectos se podrían utilizar distintas configuraciones del mismo algoritmo, resultando en distintas configuraciones y resultados para un mismo modelo de prueba. El ritmo de aprendizaje, el porcentaje de ejemplos para la validación o el tipo de algoritmo, del cual hablaremos después, son algunos más son factores que pueden influenciar los resultados y desde luego deben ser objeto de investigación y prueba para futuras investigaciones y proyectos.

Por otra parte, el algoritmo utilizado, red neuronal o KNN, puede haber afectado las decisiones tomadas por la inteligencia elegida para esta investigación. En futuras investigaciones, el algoritmo utilizado ha de ser el objetivo de futuros procesos de prueba para comprobar la probabilidad de que un tipo de algoritmo afecte a los datos obtenidos y cómo podríamos utilizar estas conclusiones para precisar y mejorar los datos, y así crear un modelo más fiable, preciso, y adaptado a las necesidades que puedan surgir en el futuro.

8. Bibliografía

Ávila Rodríguez, A. M. (2019). *Web de La Universidad de los Andes, Colombia*. Obtenido de <http://hdl.handle.net/1992/44993>

Blázquez, R. B. (2020). El sesgo de la máquina en la toma de decisiones en el proceso penal. *IUS ET SCIENTIA* (6.2), 54-71.

David Rodríguez García, J. (Mayo de 2022). *LearningML*. Obtenido de <https://web.learningml.org>

de Ágreda, Á. G., & Salazar García, I. (2019). Sesgos y perspectiva cultural en el entrenamiento de los algoritmos de inteligencia artificial. *Revista de privacidad y derecho digital* (4.15), 29-63.

- Degli-Esposti, S. (2021). El rol del análisis de género en la reducción de los sesgos algorítmicos. *Información Comercial Española, ICE: Revista de economía* (921), 127-143.
- Faliero, J. C. (2021). Limitar la dependencia algorítmica: Impactos de la inteligencia artificial y sesgos algorítmicos. *Nueva sociedad* (294), 120-129.
- Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? *Nueva Sociedad* (294), 27-36.
- López Baroni, M. J. (2019). Las narrativas de la inteligencia artificial. *Revista de Bioética y Derecho* (46), 5-28.
- López, J. M. (2022). Inteligencia artificial, sesgos y no discriminación en el ámbito de la inspección tributaria. *Crónica Tributaria* (182.1), 51-89.
- Mendoza, D. (2020). Racismo y roles de género, conductas perpetuadas en algoritmos de inteligencia artificial. *COLOQUIO* (65), 131-135.
- Rivas Vallejo, M. (2022). Sesgos de género en el uso de inteligencia artificial para la gestión de las relaciones laborales: análisis desde el derecho antidiscriminatorio. *e-Revista Internacional de la Protección Social*, VII (1), 1-30.
- Valenzuela, L. A. (2021). Los algoritmos digitales en el trabajo. Brechas y sesgos. *Relaciones Laborales y Derecho del Empleo*.