

Red neuronal para la clasificación de individuos en grupos étnicos para el censo poblacional de Colombia 2018

Abstract

En Colombia la caracterización de su población de la población es de gran ayuda para establecer la inversión de su políticas públicas, sin embargo aún no se ha desarrollado una metodología certera para obtener información real sobre algunos de sus habitantes. El desarrollo de modelos de clasificación especialmente las redes neuronales es de gran ayuda a la hora de darnos una idea sobre la posible errónea clasificación. Este trabajo desarrolla una ejecución de un modelo autoencoder para la reclasificación étnica sobre el censo Colombiano de 2018 con el cual desarrollamos una alternativa para revisar el comportamiento de los habitantes de Colombia post toma de información.

Palabras Clave: Autoencoder, encoder, clasificación, red neuronal

1.. Introducción

El Departamento Administrativo Nacional de Estadística – DANE tiene como misión fundamental garantizar la disponibilidad y la calidad de la información estadística estratégica para el desarrollo económico y político del país, en el marco de sus objetivos misionales, así como garantizar la producción, disponibilidad y calidad de la información estadística estratégica; y dirigir, planear, ejecutar, coordinar, regular y evaluar la producción y difusión de información oficial básica.

En ese sentido, el DANE tiene como objetivo mejorar la disponibilidad de estadísticas relevantes con niveles de desagregación más detallados (relacionado con grupos poblacionales, o con dominios geográficos); así como de producir estadísticas relevantes con una mayor frecuencia. Un medio para lograrlo es a partir de la integración de las diferentes fuentes de información tradicionales (como encuestas y censos) con fuentes alternativas (como imágenes satelitales, registros administrativos, técnicas de big data, entre otros).

Con el pasar de los años nos encontramos que los avances tecnológicos han hecho que podamos hacer una mejor manipulación de la información y que el procesamiento que le podamos dar a dicha información sea el mas óptimo para mejorar el conocimiento, en este caso centrado en el socio-demográfico, ayudándonos a categorizar de una manera mas exacta las diferentes poblaciones, centrándonos en los grupos étnicos, que son identificados con las preguntas de autorreconocimiento, en las diferentes encuestas que realiza periódicamente el Departamento Administrativo Nacional de Estadística (DANE), gracias a este tipo de preguntas se puede indagar a cual grupo étnico reconocido institucionalmente pertenecen en Colombia.

En el 2018 se realizó el censo de población en Colombia, el cual intenta identificar la composición de la población colombiana, esto para tomar decisiones y orientar políticas públicas en aspecto sociales, una de las variables medidas fue el reconocimiento a los diferentes grupos étnicos, debido a la posible desinformación o errónea recogida de las observaciones, un gran porcentaje de la población no se clasificó en un grupo étnico particular generando errores a la hora de identificar a la población del país. El problema fue identificado debido a que durante las proyecciones del censo del 2005 se esperaba que en el 2018 se tuviera cerca de 4 millones de afro colombianos, cifra que con el censo de 2018 fue cercana a 2 millones.

Esto se debe principalmente a dos problemas:

1. Las personas fueron mal clasificadas al momento de ser encuestadas.
2. Error por parte de los encuestadores al no tomaron el dato verídico

Para mitigar este problema principalmente se plantea un modelo de redes neuronales en que se toman a todas aquellas personas que tuvieron una errónea clasificación en su grupo étnico.

En este trabajo desarrollamos la implementación de una red neuronal autoencoder, esto para la reclasificación del grupo étnico de personas autorreconocidas como sin

ningún grupo a la categoría negros, se usaron variables sociodemográficas relacionadas con el autorreconocimiento y este método planteado se comparó con otras metodologías tales como una regresión logística y máquinas de soporte vectorial usando algunas medidas de bondad de ajuste.

2.. Antecedentes

En la planeación y el desarrollo del Censo Nacional de Población y Vivienda 2018 CNPV 2018 contaron con la participación de los grupos étnicos constitucionalmente reconocidos: indígenas, gitanos o Rrom, y comunidades negras, afrocolombianas, raizales y palenqueras. Lo anterior, de acuerdo con lo establecido en el Plan Nacional de Desarrollo 2014- 2018 “Todos por un nuevo país”, y en reconocimiento de sus derechos establecidos en el marco normativo nacional e internacional, entre otros, la Constitución Política de 1991, el Convenio 169 de 1989 sobre pueblos indígenas y tribales, la Ley 21 de 1991, la Sentencia T- 576 del 4 de agosto del 2014 y los decretos 1745 de 1995, 1066 de 2015 y 1372 de 2018.

Para la identificación de la etnia en las personas censadas se planteó mediante una pregunta en el formulario del Censo que designa directamente al encuestado si considera, o pertenece a un grupo étnico, es decir, como se autorreconoce.

En el proyecto se planteó el diseño y entrenamiento de una red neuronal para la clasificación de personas afrodescendientes, es decir, como análisis exploratorio se escoge 1 de las clases de etnia consideradas en la pregunta de autorreconocimiento étnico para evaluar el poder predictivo de la red neuronal.

En agosto de 2020, el DANE realiza por primera vez el estudio del caso, ellos proponen principalmente asignar probabilidades condicionales, dependiendo del nombre de la persona y su relación con los grupos étnicos, o residir en determinada ubicación geográfica dado que pertenece a un grupo étnico. Para el desarrollo de este estudio se plantea un modelo de clasificación de Random Forest, pero ellos solo se centraron en uno de los dos problemas propuestos inicialmente, que las personas fueron mal clasificadas al momento de ser encuestadas.

3.. Enfoque

Para el desarrollo de esta investigación se toma el censo nacional de población del año 2018 Administrativo Nacional de Estadística (DANE). Teniendo el censo como base se tomarán todas aquellas variables sociodemográficas que nos puedan brindar una caracterización de

cada uno de los individuos que contiene dicha base. Esta caracterización está centrada en el lugar donde reside, el lugar donde nacieron, sus edades, el sexo al que se identifican, y el sitio donde se encuentra la vivienda donde reside.

Para realizar la clasificación del autorreconocimiento étnico, usaremos inicialmente una red neuronal autoencoder y también una red neuronal simple, seleccionaremos a las variables sociodemográficas más relevantes y desarrollaremos una reducción dimensional de ellas, estas nuevas variables serán la capa de entrada de la red neuronal, en primer lugar tomando la red autoencoder la capa superficial está dada por la cantidad de variables que se producen al realizar un procedimiento de onehot, que lo que hace este procedimiento es crear una variable dumi de 1 y 0 para cada categoría de las variables iniciales, por ejemplo si tenemos el sexo que tiene las categorías 1 y 2, creará una variable para la categoría 1 y otra variable para la categoría 2, luego de tener nuestra nueva base, la cantidad de neuronas será igual a la cantidad de variables que tiene esta nueva variables y será utilizada para el encoder, posteriormente realizaremos el decoder para la reconstrucción del tipo del autorreconocimiento étnico, para que al ingresar un dato diferente a la etnia de interés la red lo detecte y registre.

Para el modelo propuesto revisaremos diferentes funciones de pérdida como son la entropía cruzada y el error cuadrático medio. Entrenaremos la red neuronal varias veces de tal forma que se pueda comparar diferentes redes y se puedan comparar y obtener la mejor dependiendo de la diferencia de las clasificaciones por medio de matrices de confusión y los valores de precisión de cada método.

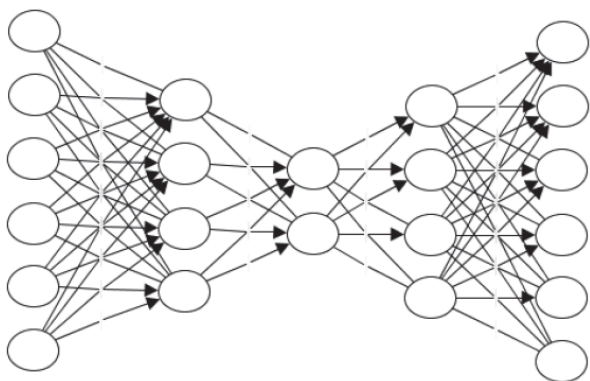
En el segundo modelo, que tomara la red neuronal simple, tendremos una modificación, y no se realizará el procedimiento onehot, si no se realizara con la base inicial, la base que se suministra del censo de población, en este caso también tendrá como referencia la cantidad de variables sociodemográficas que tiene la base, en este caso la base no será tan extensa. Luego de esto modificaremos los hiperparámetros, tales como las épocas, el tamaño de los Batch, hasta encontrar el óptimo y que nos genere un modelo con mejor predicción.

3.1.. Autoencoders

Un autoencoder es una red neuronal cuyos valores de entrada son iguales a los de salida, por lo que emplea aprendizaje no supervisado (LeCun, 1987, Ballard, 1987, Schmidhuber, 2015).

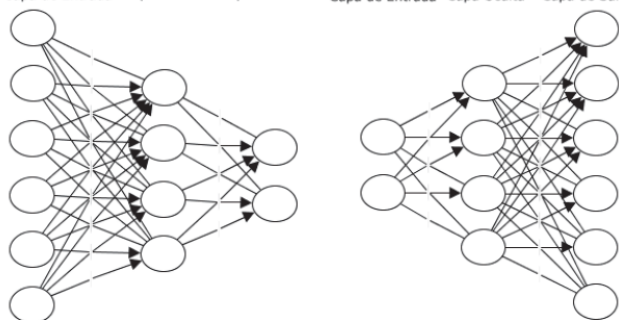
La estructura general de un autoencoder es, simétrica con respecto a la capa central y se componen de dos par-

Capa de Entrada Capa Oculta 1 Capa Central Capa Oculta 3 Capa de Salida



tes: desde la capa de entrada hasta la capa central de salida, decodificador. Esta nomenclatura se debe a que si se divide el autoencoder en dos redes neuronales funcionan como un algoritmo de codificaciones: la red neuronal cuya entrada es la entrada del autoencoder y su salida la capa central actúa como una función de comprensión de información, mientras que la red cuya capa de entrada es la capa central del autoencoder y su capa de salida la capa de salida del autoencoder actúa como una función de descompresión. El codificador funciona también como extractor de características.

Capa de Entrada Capa Oculta Capa de Salida Capa de Entrada Capa Oculta Capa de Salida



(a) Codificador.

(b) Decodificador.

Si nos fijamos, cuando se entrene la red neuronal de la figura del autoencoder anterior se calcularán los pesos de todas las conexiones de la red y con ello, podemos separar la red en 2 subredes:

Subred 1: formada por la capa de entrada y la capa oculta, el resultado obtenido de esta subred es el valor de las neuronas que forman la capa oculta, donde se habrá obtenido una transformación de las muestras reales a un espacio alternativo. Por esta razón, a esta subred se le llama encoder

Subred 2: formada por la capa oculta y la capa de salida, en este caso, partiendo del resultado obtenido en las neuronas de la capa oculta en el paso anterior (resultado del encoder) se obtiene el valor de las neuronas de la capa de salida, que recordemos, sus pesos habían sido entrenados (junto con los demás de la red) para reconstruir la entrada de salida. Con ello, podemos obtener a partir del resultado del encoder el resultado original. Por esta razón, a esta subred se le llama decoder

Por tanto, como se ha visto, al ser el autoencoder simple un caso particular de red neuronal, podemos hacer uso del script implementado en el apartado anterior, que nos permite diseñar redes neuronales y aprender los pesos que minimicen el error cuadrático medio. Consideraré 2 ejemplos de uso de estos autoencoder, el cifrado de bloques de texto y la compresión.

Los autoencoders se emplean típicamente para la reducción de dimensionalidad de datos (Hinton and Salakhutdinov, 2006), la extracción de características (LeCun, 1987) o en aprendizaje por refuerzo (Lange and Riedmiller, 2010) entre otros.

3.2.. Red neuronal densa.

Es el tema con mayor cantidad de investigaciones en el aprendizaje automático, método conocido como perceptrón multicapa, ha sido diseñado para simular las operaciones del sistema nervioso humano, la forma más simple de una red neuronal es un solo perceptrón. El elemento esencial de un perceptrón son los valores de entrada, con pesos asociados, sesgo y funciones de activación que generan un resultado. Las funciones de activación más usadas son el sigmoide, tangente hiperbólica (Tanh) y (ReLU). Una sofisticada red neuronal es una herramienta universal para simular cualquier función y obtener un excelente nivel de accuracy [15].

A nivel esquemático una neurona artificial se representa así:

Esta es una red neuronal de un solo perceptrón, compuesta por sus valores de entrada, pesos y sesgo junto con su propia función de activación que genera una respuesta de salida.

Esquema de una red neuronal densa o “perceptrón multicapa”:

Las neuronas de la primera capa reciben como entrada los datos reales que alimentan a la red neuronal. Es por eso por lo que la primera capa se conoce como capa de entrada. La salida de la última capa es el resultado visible

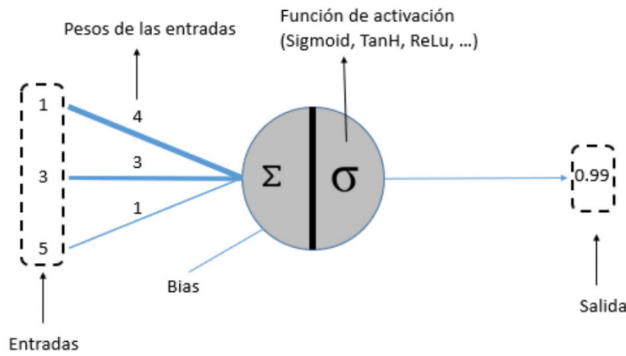


Figura 1. Esquema de una neurona

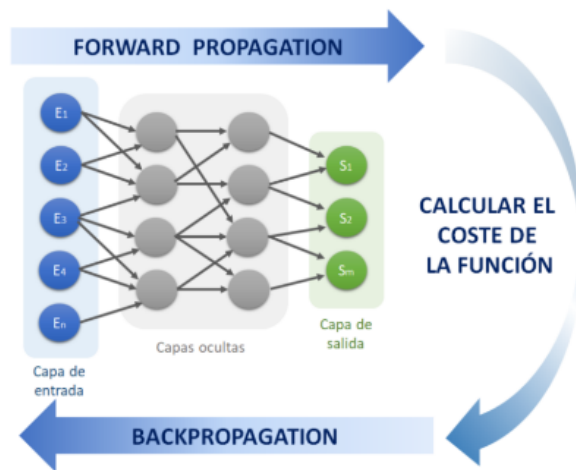


Figura 2. Esquema de una red neuronal densa

de la red, por lo que la última capa se conoce como la capa de salida. Las capas que se sitúan entre la capa de entrada y la capa de salida se conocen como capas ocultas ya que desconocemos tanto los valores de entrada como los de salida.

Parámetros e hiperparámetros:

Los parámetros se calculan de manera iterativa a través del entrenamiento de la red mientras que los hiperparámetros son configurados de manera discrecional por el diseñador de la red. Existen hiperparámetros tanto a nivel de estructura de la red como a nivel de algoritmo de aprendizaje. A continuación, se hace mención de los hiperparámetros más importantes:

- **Épocas (epochs):** Indica el número de veces que todos los datos pasan por la red neuronal en su etapa de entrenamiento para el ajuste de los parámetros.
- **Batch size:** los datos de entrenamiento se pueden di-

vidir en mini lotes, estos son subconjuntos de datos que se seleccionan aleatoriamente para la actualización del gradiente que disminuye la función de costo, dependiendo del número de subconjuntos que se elijan respecto al total de datos será el número de actualizaciones que se llevarán a cabo en una época.

- **Learning rate:** este hiperparámetro es el salto que debe dar el algoritmo en la búsqueda del mínimo de la función de costo. Este valor mide cuánto varían los parámetros en cada actualización. Un valor alto indica saltos grandes, es decir, el aprendizaje será rápido, sin embargo, se puede saltar el mínimo de la función de costo o incluso la red podría llegar a divergir. Si el learning rate es pequeño se tendrá más posibilidades de encontrar el mínimo, pero el aprendizaje será lento. Existe la posibilidad de modificar este hiperparámetro en función de los resultados obtenidos durante el entrenamiento.

Proceso de aprendizaje de una red neuronal:

- **Forward propagation:** Este proceso que parte de las neuronas de la capa de entrada y termina en las neuronas de la capa de salida, durante este procedimiento cada neurona realiza una suma ponderada de todas las entradas de acuerdo a unos pesos, pasa el resultado por una función de activación y genera el resultado, que pasa a la siguiente capa, esto se repite hasta la última capa de neuronas que nos devuelve el resultado de la red.

- **Cálculo de la función costo:** La función de costo calcula el error entre el valor estimado por la red y el valor real, con el fin de optimizar los parámetros de la red neuronal.

- **Backpropagation:** Finalmente, a partir del resultado de la función de costo el algoritmo propaga la información de esta función a través de las neuronas de la capa oculta y se actualizan los parámetros utilizando algún método de optimización, por el método descenso del gradiente, inmediatamente se calcula la derivada de la función de costo y se trata de avanzar hacia el mínimo global.

- **Función de activación:** Es la función que se aplica al sesgo más la suma de las entradas multiplicadas por los pesos. Esta función simula el umbral de activación en una neurona biológica. Algunas de las funciones más utilizadas son; Relu, Sigmoide y Tangente hiperbólica.

Una red neuronal, por lo tanto, siempre está compuesta por una capa de entrada, una capa de salida (si solo hay una

capa en la red neuronal, la capa de entrada coincide con la capa de salida) y puede contener 0 o más capas ocultas. El concepto de Deep Learning nace a raíz de utilizar un gran número de capas ocultas en las redes [16].

Entrenar una red neuronal consiste en ajustar cada uno de los pesos de las entradas de todas las neuronas, para que las respuestas de la capa de salida se ajusten lo más posible a los datos que conocemos, para el caso de clasificación tendríamos una neurona con una función de activación sigmoide que daría como resultado 1 en la predicción del individuo afrodescendiente y 0 si no lo es.

4.. Metodología

Para el desarrollo de este proyecto, se toma el censo poblacional realizado por el DANE, solo se utilizará la base a nivel de personas, ya que esta base tiene toda información en la cual se centra nuestro estudio, cabe resaltar que existe una base a nivel de vivienda y otra a base de hogar, pero estas dos bases no serán requeridas en este estudio.

Basados en la base poblacional, seleccionamos algunas variables socio económicas, estas variables que se seleccionan están relacionadas con el lugar de procedencia, el lugar de donde viven, la pertenencia étnica, etc. Teniendo en cuenta las variables se entrena en primer lugar la red neuronal autoencoder, teniendo en cuenta diferentes tamaños lotes("Batch"), los tamaños de lotes nos ayuda a que tan rápido aprende la red, es decir, los lotes hacen tienden a decirnos de cuantos grupos de registros entran en la red para enseñarla, por tanto hicimos ajustamos que el tamaño del lote sea lo suficientemente grande y pero que a su vez proporcione la información necesaria.

No solo se hizo una variación en los lotes, ya que las redes neuronales también están constituidas con mas hiperparametros, entonces también tenemos que tener en cuenta la función de activación, entrenamos la red con una función de activación RELU, función tangencial, sigmoideal y "softmax" dependiendo del comportamiento de los datos seleccionaremos la mejor función que haga aumentar la precisión de la predicción en la red neuronal.

Los pasos anteriores se realizaron tanto con la red neuronal Auto Encoder como la red neuronal simple, de tal forma que se pueda hacer una comparación entre estas dos redes y de esta manera entrenar con el mejor red neuronal, y así hacer una predicción y la precisión de sea lo mejor posible, pero sin tener problemas de sobre ajuste.

5.. Resultados

Teniendo en cuenta que estamos realizando los todo el procedimiento del censo nacional de población, y conocemos la distribución poblacional del país nos centraremos directamente en el entrenamiento y testeo de las redes neuronales:

5.1.. Resultados Autoencode

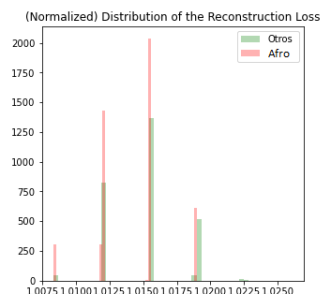
5.1.1. Primera red

En esta primera red tenemos una función de activación relu, con 100 épocas y 100 lotes, con 1141 neuronas en la capa superior, y en las capas ocultas tenemos 500, 300 y 10 neuronas. estos hiperparametros fueron escogidos ya que después de haber realizado el procedimiento muchas veces fueron los valores que tuvieron una muy buena precisión.

Las variables que utilizamos fueron las siguientes:

- P_SEXO: Sexo.
- PA11_COD_ETNIA: Pueblo indígena de pertenencia.
- PA12_CLAN: Clan de pertenencia.
- PA_LUG_NAC: Lugar de nacimiento.
- PA1_DPTO_NAC: Departamento de nacimiento.
- P_PARENTESCO: Relación de parentesco con el jefe(a) del hogar.
- UA_CLASE: Clase.
- UVA_ESTATER: Vivienda en una territorialidad étnicaGMO.
- UVA1_TIPOTER: Tipo de territorialidad étnica.
- UVA2_CODTER: Código de territorialidad étnica.
- PA1_GRP_ETNIC: Reconocimiento étnico – Variable de respuesta.

Luego de entrenar la red neuronal auto encoder y realizar la reconstrucción obtenemos lo siguiente:



Como se observa en la gráfica puede parecer que la reconstrucción es muy buena, que estaríamos encontrando todas aquellas personas que se identificaron como afros durante el censo de poblacional, sin embargo cuando miramos que tan preciso fue el modelo encontramos los siguientes resultados:

```
The classifications using the MAD method with threshold=1 are as follows:
[[3014852  76224]
 [ 910237  63287]]

% of transactions labeled as fraud that were correct (precision): 63287/(76224+63287) = 45.36%
% of fraudulent transactions were caught successfully (recall): 63287/(910237+63287) = 6.58%
```

como se observa en este primer modelo se obtiene una precisión del 45 % lo cual no nos pareció la mejor opción, también nos damos cuenta que según la matriz de confusión no está clasificando bien, ya que tiene un recall de 6.5 %, es decir que clasifica a menos personas afros cuando en realidad son afros, lo cual hace que descartemos este modelo y continuemos viendo para el segundo modelo.

5.1.2. Red neuronal densa.

Una alternativa ha sido entrenar una red neuronal densa para la predicción del autoreconocimiento étnico específicamente la categoría afrodescendiente, la estructura de la red está compuesta por 1138 neuronas de entrada, dos capas intermedias de 600 neuronas y una capa de 2 neuronas de salida que permitirá identificar la respuesta binaria de la predicción, donde 1 el individuo fue clasificado como afrodescendiente y 0 no afrodescendiente.

Las variables usadas para el aprendizaje de la red neuronal han sido:

- P_SEXO: Sexo.
- PA11_COD_ETNIA: Pueblo indígena de pertenencia.
- PA12_CLAN: Clan de pertenencia.
- PA_LUG_NAC: Lugar de nacimiento.
- PA1_DPTO_NAC: Departamento de nacimiento.
- P_PARENTESCO: Relación de parentesco con el jefe(a) del hogar.
- UA_CLASE: Clase.
- UVA_ESTATER: Vivienda en una territorialidad étnicaGMO.
- UVA1_TIPOTER: Tipo de territorialidad étnica.
- UVA2_CODTER: Código de territorialidad étnica.
- PA1_GRP_ETNIC: Reconocimiento étnico – Variable de respuesta.

Para balancear las categorías de la variable referente al autoreconocimiento étnico se creó un nuevo conjunto de datos que contiene los 2'950.072 individuos que se auto reconocieron como afrodescendientes y con un muestreo aleatorio simple se extrajeron 2'961.186 individuos que no se auto reconocieron como afrodescendientes. Finalmente se obtiene un conjunto de datos con una participación de 49,9 % para la categoría afrodescendiente y 50,1 % para los no afrodescendientes, con un total de 5'911.258 individuos pertenecientes a la base del censo de población y vivienda 2018.

Posteriormente se discretizan las variables categóricas y se genera un conjunto de datos de entrenamiento y de evaluación del conjunto de datos que corresponden al 70 % y 30 % respectivamente. Finalmente se entrena el modelo. A continuación, se presenta la tabla de clasificación con las métricas de evaluación.

	precision	recall	f1-score	support
0	0.59	0.46	0.52	888573
1	0.56	0.69	0.62	885429
accuracy			0.57	1774002
macro avg	0.58	0.57	0.57	1774002
weighted avg	0.58	0.57	0.57	1774002

Figura 3. Tabla de clasificación red neuronal densa

En la evaluación del modelo se obtiene un “accuracy” del 57 % (porcentaje total de aciertos en el modelo), se debe resaltar que el modelo presenta una sensibilidad del 69 %, es decir, logra identificar el 69 % de los individuos que se auto reconocieron como afrodescendientes en el conjunto de datos de evaluación.

	pred_noafro	pred_afro
R_noafro	405166	483407
R_afro	276683	608746

Figura 4. Matriz de confusión
x

Al observar la tabla 2 el modelo identifica ?? 608.746 individuos que se reconocieron como afrodescendientes frente a un total de 885.429 afrodescendientes que contenía el conjunto de datos de evaluación, sin embargo, hay 483.407 individuos que el modelo clasifica de manera errónea afirmando que son afrodescendientes, de aquí se identifica el 62 % de los casos identificados por el modelo corresponden a la realidad.

Metrica AUC para evaluar el modelo red densa: 0.5717445545599583

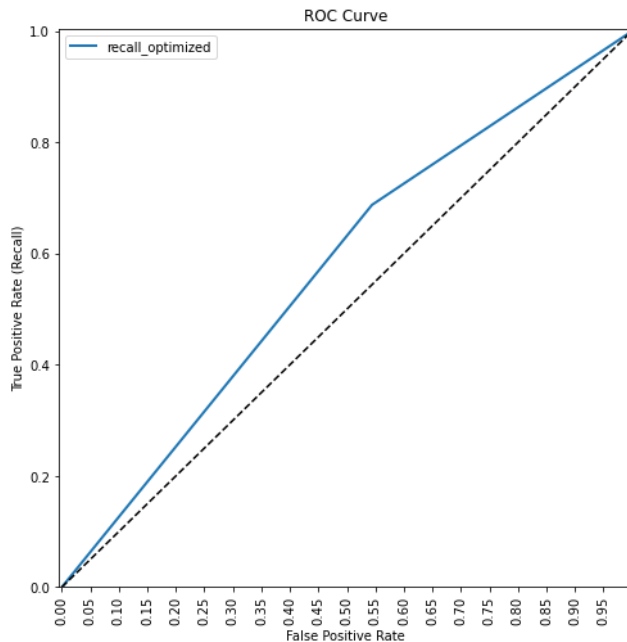


Figura 5. Curva ROC red neuronal densa

La Figura 5 es una prueba que permite determinar la capacidad discriminativa del clasificador según el área bajo la curva; mientras mayor sea más poder discriminatorio tendrá el clasificador. Esta área puede interpretarse como la probabilidad de que, ante un par de individuos, uno que se auto reconozca afrodescendiente y el otro no, el modelo los clasifique correctamente.

6.. Conclusiones

La revisión de las funciones de activación dentro del autoencoder, muestran que la función de activación sigmoideal tiene mejor comportamiento en medidas como accuracy, precisión y Recall, esto debido a que se incluyeron solamente variables cualitativas para entrenar la red neuronal ajustándose como vectores one-hot y nuestra respuesta para clasificar era una variable dicotómica.

La clasificación realizada por la red neuronal puede mejorar, teniendo en cuenta que se omitieron algunas variables que pueden llegar a categorizar el autorreconocimiento étnico con mayor precisión, estas variables pueden ser las que estén relacionadas con la educación de cada uno de los individuos encuestados, metodologías como las maquinas de soporte vectorial y la regresión logística tienen mejor comportamiento en el accuracy que nuestro modelo.

Para mejorar la red neuronal, se debe seleccionar una

muestra, en donde se debe tener en cuenta una estratificación de acuerdo, a la variabilidad étnica por departamento, esto en el caso de que no se tenga la potencia computacional para entrenar y procesar la red neuronal.

7.. Referencias

- Le Cun, Y.(1987) modeles connexionnistes de l'apprentissage. PHD thesis, PHD thesis, These de Doctorat, Universite de paris.
- Schmidhuber, j. (2015). Deep learning in neural networks: An overview. Neural Networks, 61:85 -117
- Lange, S. and Riedmiller, M. (2010) Deep autoencoder neural network in reinforcement learning. In Neural Networks (IJCNN), the 2010 International Joint Conference on, pages 1-8 IEEE.
- Vicent, P., Larachelle, H., Bengio, Y., and Manzagol, P. -A. (2008) Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine Learning, Pages 1096 -1103 ACM.
- Le Cun, Y., Wan, L., Zeiler, M., Zhang S., and Fergus, R. (2013). Regularization of neural networks using dropconnect. In International Conference on Machine Learning, pages 1052 - 1066.
- Minsky, M. and Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge, MA, USA.
- <http://microdatos.dane.gov.co/index.php/catalog/MICRODATOS>
- Departamento Administrativo Nacional de Estadística-DANE-. (2018). Informe comité de expertos internacional
- Departamento Administrativo Nacional de Estadística-DANE-. (2018). Informe comité de expertos nacional