

Resumen

El proyecto tiene como propósito la predicción del genero de una persona mediante una grabación de voz.

Se usaron varios clasificadores (naive bayes classifier, Support vector machine, decisión tree classifier y Random forest classifier) y además de redes neuronales poder hacer la clasificación del genero.

Introducción

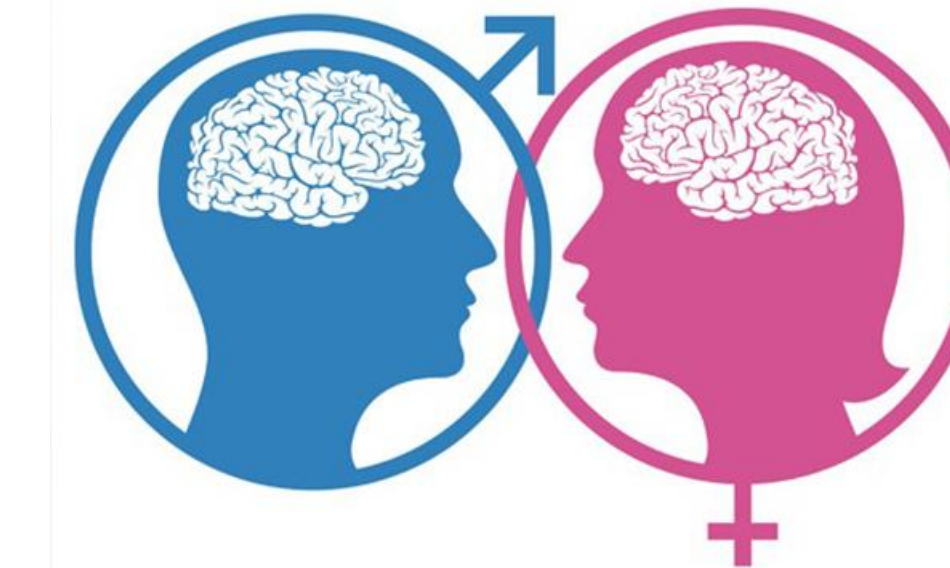
El propósito de este proyecto fue crear una inteligencia artificial que haga reconocimiento de voz y a partir de este registro identificar el genero biológico del usuario, esta inteligencia va a ser entrenada para poder realizar dicho reconocimiento, también se entrenaron 2 redes neuronales, se usa un csv anteriormente descargado.

Se utilizará un dataset que contiene las propiedades acústicas de la voz y speech, el dataset está compuesto de 3,168 muestras de voz, recolectadas de hombres y mujeres. Las muestras de voz fueron pre-procesadas mediante un análisis acústico en R usando los paquetes seewave y TuneR, con un rango de frecuencia de 0hz-280hz.(Rango de voz humana). Se entrenaran 4 clasificadores de machine learning (Naive Bayes classifier, Support vector machine, Decision tree y Random Forest) y se determinara cual obtiene el mejor puntaje de clasificacion y por lo tanto mejor clasificacion del dataset. También se entraron 2 redes neuronales para poder clasificar, una red con más capas y con menos epochs (numero de iteraciones) mientras que la otra con menos capas y más epochs .

Proceso y método

El proyecto tiene las siguientes características:

- Se usa el dataset de las grabaciones de las personas las cuales se clasifican como hombre o mujer
- El dataset contiene 3168 muestras de voz en las cuales la mitad son hombres y la otra mitad mujeres
- Las muestras fueron pre-procesadas mediante un análisis acústico en R usando los paquetes seewave y TuneR, con un rango de frecuencia de 0hz-280hz.



Resultados

En este estudio al comparar los resultados de los cuatro clasificadores y las dos redes neuronales, pudimos comprobar que la máquina de soporte vectorial y Random Forest obtienen los mejores resultados entre clasificadores con 98.7% y 98.4% respectivamente, y la red neuronal con 3 capas y 50 epochs obtiene un puntaje superior con 98.8% en comparación a la red neuronal de 6 capas y 20 epochs.

Sin embargo, el clasificador con menor puntaje que en este caso es Naive Bayes, obtuvo 93.2% de precisión que sigue siendo un buen puntaje, pero está bastante alejado de la media en comparación al resto de puntajes, ya que tomando en cuenta los 4 clasificadores y las 2 redes neuronales se logra una media de 97,47%. Se obtuvieron estos resultados tan acertados debido a el balance del dataset y el preprocesamiento de los datos, ya que adicionalmente estandarizamos los datos y seleccionamos las características más importantes para evitar redundancias.

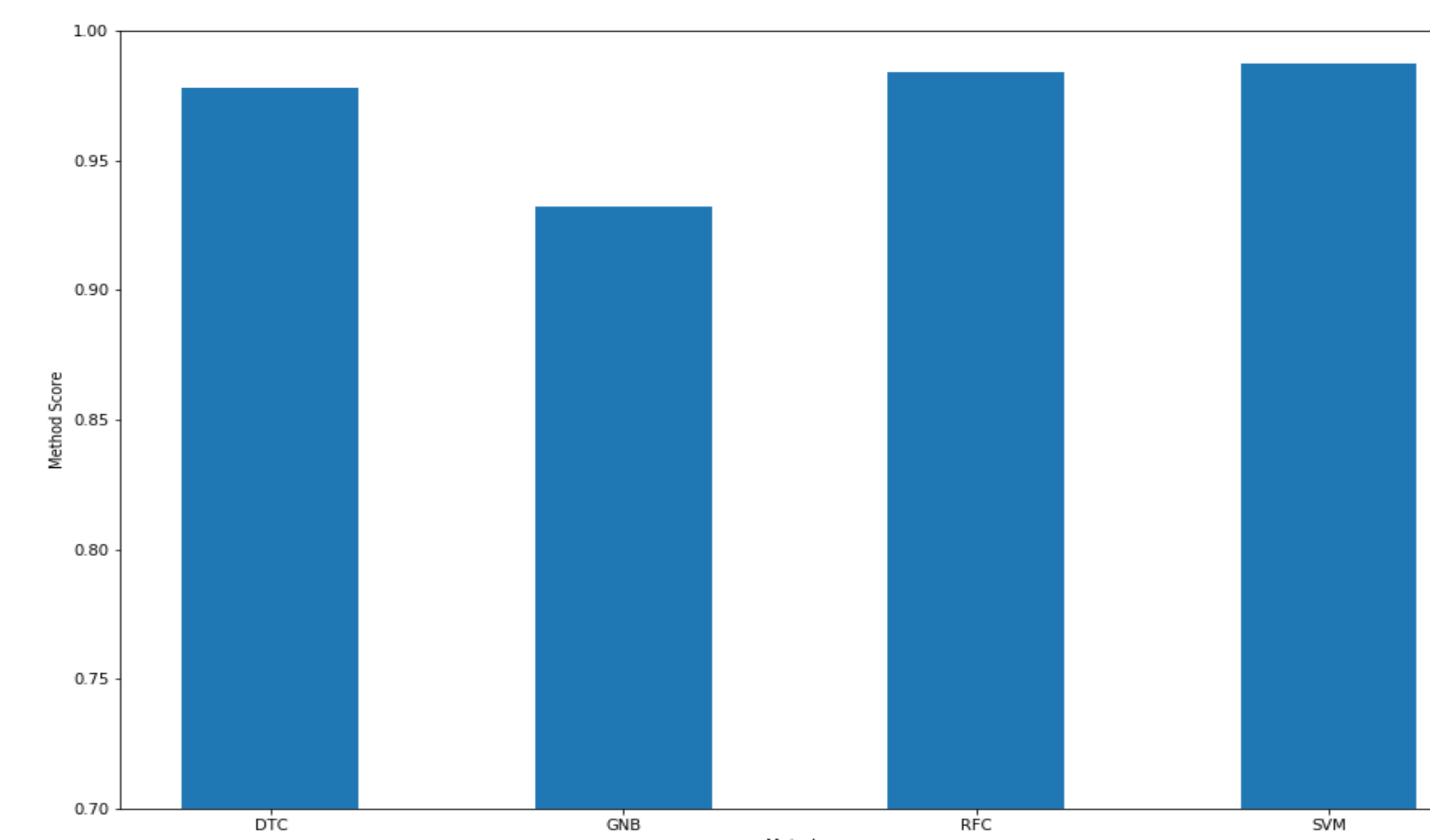
En la tabla 1 se puede observar con precisión el puntaje obtenido por cada clasificador y las redes neuronales.

En el grafico 1 se puede comparar de manera gráfica el resultado de los clasificadores en este problema.

Tabla 1. Puntajes Clasificadores

Clasificador	Puntaje
Decision Tree	0.977917981073
SVM	0.98738170347
Random Forest	0.984227129338
Naïve Bayes	0.932176656151
NN 6 layers	0.978706623477
NN 3 layers	0.988170345874

Grafico 1. Comparación entre clasificadores



Conclusiones

En este proyecto podemos concluir que tanto las redes neuronales como los clasificadores son viables ya que tienen un comportamiento muy similar y por lo tanto resultados similares.

También podemos concluir que tanto los clasificadores y las redes neuronales obtuvieron resultados muy elevados debido a que el dataset está perfectamente balanceado. Es decir, exactamente la mitad del dataset es de una clase y la otra mitad de la otra, además de que se realizó preprocesamiento de los datos.

Teniendo esto en cuenta. La clasificación con mayor puntaje y por lo tanto más idóneo para trabajar con este dataset es la red neuronal con 3 capas y 50 epochs, esta red neuronal cuenta con una capa de entrada, una capa densa de 1000 neuronas con la función de activación relu y una capa de salida con la función de activación sigmoide y se entrenó con 50 epochs. Obteniendo un puntaje de 98.81%.

La segunda alternativa es trabajar el dataset con SVM ya que logro un puntaje de 98.73% que es el segundo mejor puntaje logrado en este trabajo.

Trabajo Futuro

Desarrollar formas de hacer el producto eficiente en el uso practico, proponiendo que el usuario tenga que decir una frase especifica para que el programa tenga que analizar en la base de datos esta frase y así se limite el análisis

Información de contacto

Nombre Nicolas Galvan Alvarez, Email: nicogalvan1@hotmail.com
Nombre Hazel David Pinzón Uribe, Email: hazelpinzon13@gmail.com
Nombre Mateo Orozco Ardila, Email: oromateo@hotmail.com

Docente: Gustavo Garzón, gustavo.garzon@saber.uis.edu.co

Referencias Bibliográficas (en formato APA)

1. José Luis, Sergio Suárez. (2006). Algoritmos y Métodos para el Reconocimiento de Voz en Español Mediante Silabas. Computación y Sistemas Vol 9 Núm 3 pp 270-286. Recuperado en 06 de septiembre de 2020, de <http://www.scielo.org.mx/pdf/cys/v9n3/v9n3a7.pdf>.
2. Fosler et al. (1999). Fosler-Lussier E., Greenberg S., Morgan N., "Incorporating Contextual Phonetics into Automatic Speech Recognition". XIV International Congress of Phonetic Sciences, pp. 611-614, San Francisco, 1999.
3. Kirschning (1998). Kirschning Albers Ingrid, "Automatic Speech Recognition with the Parallel Cascade Neural Network", PhD Thesis, Tokyo Japan, March 1998.
4. Kosko (1992). Kosko B., "Neural Networks for Signal Processing", Prentice Hall, U.S.A., 1992. . Sydral et al. (1995). Sydral A., Bennet R., Greenspan S., "Applied Speech Technology", Eds (1995). CRC Press, ISBN 0-8493-9456-2, U.S.A., 1995.