

Sistema de Clasificación de Géneros Musicales para plataformas de Streaming mediante Algoritmos Genéticos

Una metaheurística inspirada en los principios de
la selección natural

Jorge Eduardo Enciso Agudelo, Daniel Figueroa Torres, Sergio Andrés Rojas Moreno,
Edison Javier Díaz Prieto, Alexander Luna Ruiz, Árnol Guerrero Calderón

to

2. Objetivo

4. Descripción de la Tarea

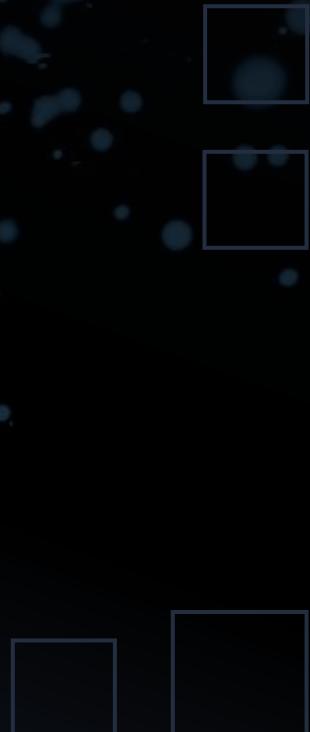


3. Estado del Arte

5. Análisis de Viabilidad



Contexto



Contex^to

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Dentro de los retos que surgieron de la primera entrega del proyecto del “Sistema de Clasificación de Géneros Musicales para Plataformas de Streaming”, se encontraba el poder encontrar el clasificador más adecuado para mejorar las métricas de desempeño que se obtuvieron inicialmente. Gracias a los Algoritmos Genéticos por medio de Python, se podrán maximizar dichas métricas de desempeño, por lo tanto, podemos aproximarnos a mejores resultados.

Con esto es posible continuar con el problema inicial de clasificación, pues con mejores métricas se podrá crear un nuevo sistema con la capacidad de segmentar de manera más eficiente los géneros musicales



Contexto

Es difícil asegurar que con todos los géneros se va a obtener un resultado muy específico de clasificación, por el contrario, se desea llegar a un punto donde se pueda dar un porcentaje de similitud a los diferentes géneros, pues existen muchos géneros basados en otros, por lo que al final, lograr encontrar dicha mezcla sería más eficiente para el proceso de clasificación.

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



to

2. Objetivo

4. Descripción de la Tarea



3. Estado del Arte

5. Análisis de Viabilidad

Objetivos

1

Usar una herramienta que permita implementar Algoritmos Genéticos para mejorar el resultado obtenido en la primera entrega del proyecto, donde se pretendió clasificar los diferentes géneros musicales basado en características medidas para cada canción.

Maximizar mediante algoritmos genéticos las métricas de desempeño del algoritmo, con el fin de obtener mejores resultados a los ya obtenidos con la aplicación de Redes Neuronales Artificiales

2

to

2. Objetivo

4. Descripción de la Tarea

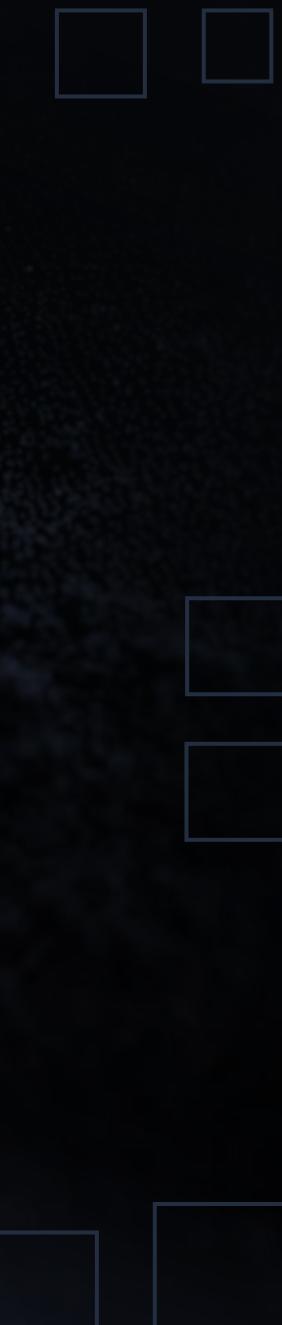


3. Estado del Arte

5. Análisis de Viabilidad

Una metaheurística inspirada en los principios de la selección natural

Estado del Arte



Automatic Music Genre Classification Using Hybrid Genetic Algorithms

George V. Karkavitsas¹ and George A. Tsihrintzis²

Advantage S.E

¹ Financial System Experts, Alimos 174 55, Greece
George.Karkavitsas@fiserv.com

² University of Piraeus, Department of Informatics
Piraeus 185 34, Greece
geoatsi@unipi.gr



Estado del Arte

Puntos Claves

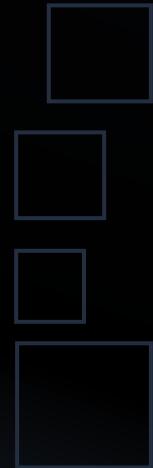
1

Desarrollar un sistema de clasificación automática de género musical y se centra en el cálculo de algoritmos que pueden predecir la clase de música a la que pertenece un archivo de música

1

2

Se utilizan mil archivos de música para entrenar y validar el sistema de clasificación. Estos archivos se distribuyen por igual en diez clases



Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Una metaheurística inspirada en los principios de la selección natural

Estado del Arte

Puntos Claves

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

De cada archivo, se extraen ochenta y un (81) características y se utilizan para crear 81 matrices de similitud. Estas 81 matrices de similitud constituyen las instancias de entrenamiento.

3

4

Durante la fase de entrenamiento, la selección de características se lleva a cabo mediante un Algoritmo Genético Híbrido modificado para mejorar la claridad de discriminación de clases y reducir el costo de cálculo



Una metaheurística inspirada en los principios de la selección natural

Estado del Arte

Puntos Claves

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

5

Se utiliza una mutación adaptativa, una búsqueda exhaustiva local adaptativa y una estrategia de reemplazo adaptativo, dependiendo de si el sistema ha llegado a un extremo local

6

La búsqueda local exhaustiva se lleva a cabo en los cromosomas vecinos óptimos hasta la solución actual. La función de adecuación del Algoritmo Genético constituye un clasificador ponderado de vecinos más cercanos. Por lo tanto, la aptitud del cromosoma es proporcional a la precisión del clasificador que crea el cromosoma



Una metaheurística inspirada en los principios de la selección natural

Estado del Arte

Puntos Claves

7

Durante la fase de clasificación, las características seleccionadas a través del algoritmo genético crean un clasificador vecino más cercano ajustado que realiza las clasificaciones.

De cada nuevo archivo de música pendiente de clasificación, las características seleccionadas se extraen y luego se comparan con las características correspondientes de los archivos de música de la base de datos. El archivo de música se asigna a la clase indicada por k archivos de música más cercanos.

8

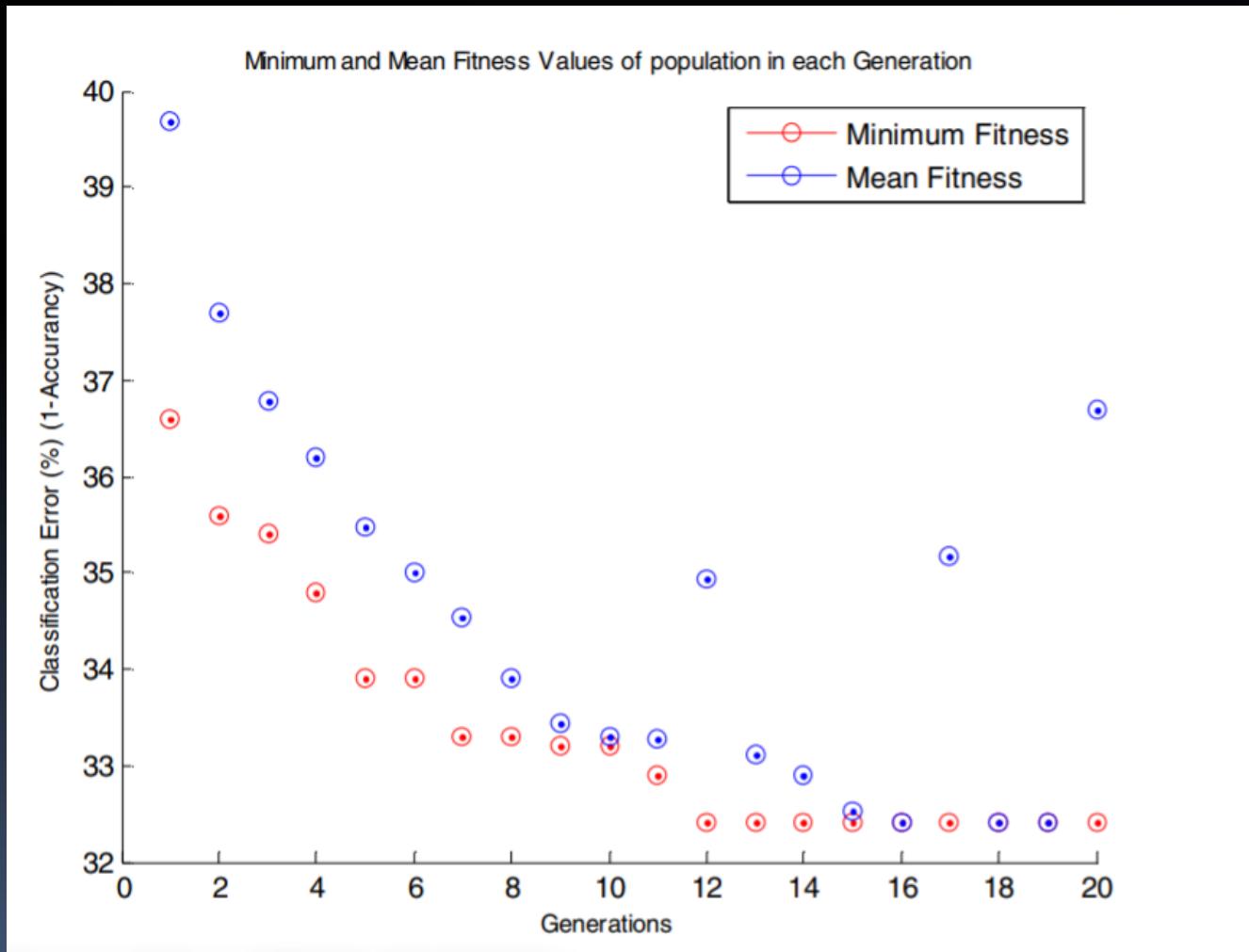
Sistema de Clasificación de Músicos Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



Estado del Arte

Puntos Claves

Una metaheurística inspirada en los principios de la selección natural



Mejor aptitud cromosómica en cada generación y aptitud media de todos los cromosomas en cada generación

Sistema de Clasificación de Géneros Musicales para plataformas de Streaming mediante Algoritmos Genéticos

(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 9, No. 12, 2018

Hyper Parameter Optimization using Genetic Algorithm on Machine Learning Methods for Online News Popularity Prediction

Ananto Setyo Wicaksono¹, Ahmad Afif Supianto²

Department of Informatics
Faculty of Computer Science, Brawijaya University
Malang, Indonesia



Estado del Arte

Puntos Claves

Se busca identificar la noticia con mayor probabilidad de tener popularidad, para que así los consumidores de noticias se vean atraídos.

1

2

Para ello, en el artículo se explican diferentes métodos de aprendizaje de máquina, donde, sin embargo, se debieron optimizar los hiper parámetros de método aplicado, para determinar cuál es el que tiene la mejor precisión.

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



Estado del Arte

Puntos Claves

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Para realizar esta optimización, los autores deciden utilizar el método de algoritmo genético con la finalidad de encontrar los mejores hiper parámetros de cada método en el menor tiempo posible. Para esto, primero se debe definir cuales hiper parámetros van a ser modificados para cada modelo

3

4

Una vez definidos, luego se define la tasa de cruce, la tasa de mutación, el número de iteraciones, el tamaño de la población y el número de generación. Al correr el método se puede identificar que le tomaba menor tiempo posible optimizar los hiperparámetros comparado con grid search.

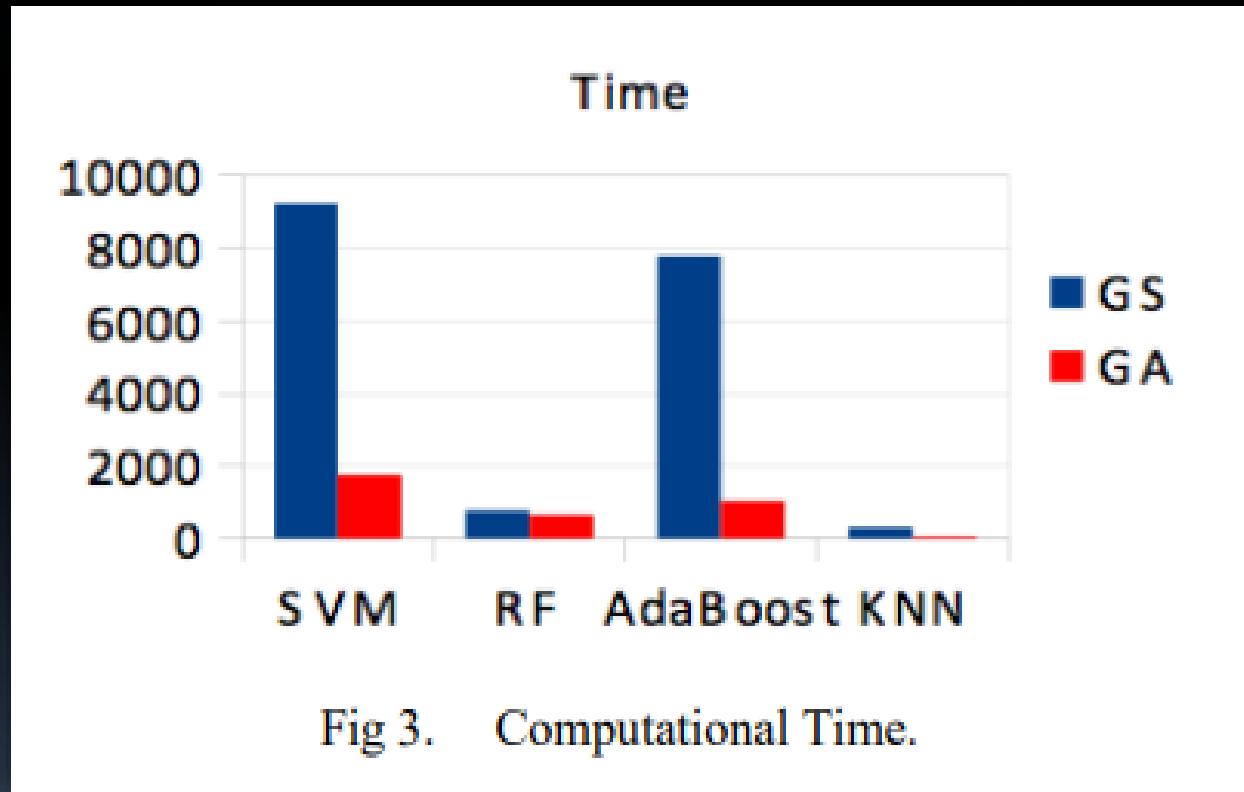


Una metaheurística inspirada en los principios de la selección natural

Estado del Arte

Puntos Claves

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



Estado del Arte

Puntos Claves

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

TABLE I. HYPER PARAMETER GRID SEARCH RESULT

Methods	Grid Search	
	Accuracy	Time
SVM	0.53	9241
RF	0.67	771
AdaBoost	0.66	7796
KNN	0.58	303

TABLE II. HYPER PARAMETER GENETIC ALGORITHM RESULT

Methods	Genetic Algorithm	
	Accuracy	Time
SVM	0.54	1760
RF	0.67	659
AdaBoost	0.66	1038
KNN	0.58	61



to

2. Objetivo

4. Descripción de la Tarea

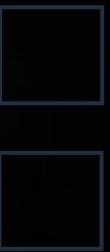


3. Estado del Arte

5. Análisis de Viabilidad



Descripción de la tarea



Descripción de la tarea

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

- Se usa un Algoritmo Genético para definir el clasificador que maximice las métricas de desempeño del algoritmo. Se parte de un modelo base, el cual fue Naive Bayes con un valor de 0.33 pero el mejor modelo fue gracias al Random Forrest con tfidf en el título de las canciones y un valor de 0.57.

Se utiliza la herramienta tpot, exactamente el TPOTClassifier para ejecutar el algoritmo genético haciendo varios intentos con diferentes valores de generaciones y de tamaño de población.



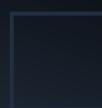
Descripción de la tarea

Restricciones

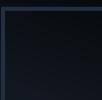
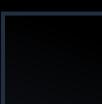
Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



Las restricciones con la aplicación de los GA en este problema sería que los individuos o soluciones para el modelo lleguen a ser factibles. Es decir que a través del crossover y la mutación la factibilidad se tiene que mantener en todo momento



Inicialmente, se requiere hacer un balanceo de los datos, pues de las diferentes clases se ven unos porcentajes desproporcionados que producen fallas a la hora de querer entrenar la red.

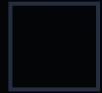


Descripción de la tarea

Restricciones

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Para los casos de la música alternativa y lo que encierra el género indie, hemos encontrado que dentro de ese mismo género siempre van a existir subgéneros muy marcados, por lo tanto, si queremos clasificar un género tan variante, va a ser muy complicado lograr una exactitud del modelo que sea significativa dado que existen otros géneros que presentan características mucho más evidentes.

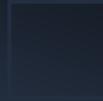


Descripción de la tarea

Alcance



El alcance para esta aplicación de AG será una búsqueda más amplia de Algoritmos de Machine learning que permitan encontrar unos parámetros que puedan minimizar el error.



to

2. Objetivo

4. Descripción de la Tarea



3. Estado del Arte

5. Análisis de Viabilidad



Análisis de Viabilidad



Análisis de Viabilidad

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

- Se describen los métodos empleados para el tratamiento de los datos, preprocessamiento, selección de características, aplicación de algoritmos de machine learning maximizando la métrica de precisión usando Algoritmos Genéticos y validación en el proceso de clasificación de los géneros musicales de acuerdo con las características escogidas.

Análisis de Viabilidad

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

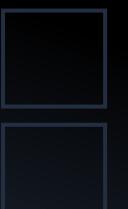
Selección de la Base de Datos

Para la selección de la base de datos, se tomó un dataset de un concurso de la página de machinehack.com en asociación de la plataforma de concursos Kaggle. El dataset es un extracto de las canciones más famosas en el año 2020 que se tomó como referencia para el concurso.



Para la limpieza de los datos se realizó una exploración de datos, en donde tenía el objetivo de encontrar datos atípicos o revisar los datos faltantes de los datos. En primera instancia, se identificó alrededor del 5% de datos faltantes en la variable del ranking de popularidad y en la variable de instrumentalidad. .

Preprocesamiento de los datos



Análisis de Viabilidad

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



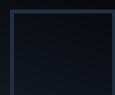
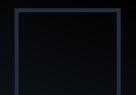
Imputación de Valores

Debido a que el porcentaje de faltantes no era significativamente alto, se decidió en imputar los valores. En este caso se usó un Random Forest con entrada las otras variables para poder predecir los registros faltantes de las dos variables



Estandarización de las variables

Con el fin de poder usar los algoritmos de Machine Learning y de Deep Learning se decidió en estandarizar todas las variables cuantitativas usando la inversa de una función uniforme, o como bien se conoce, usando el algoritmo de min max.



Análisis de Viabilidad

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



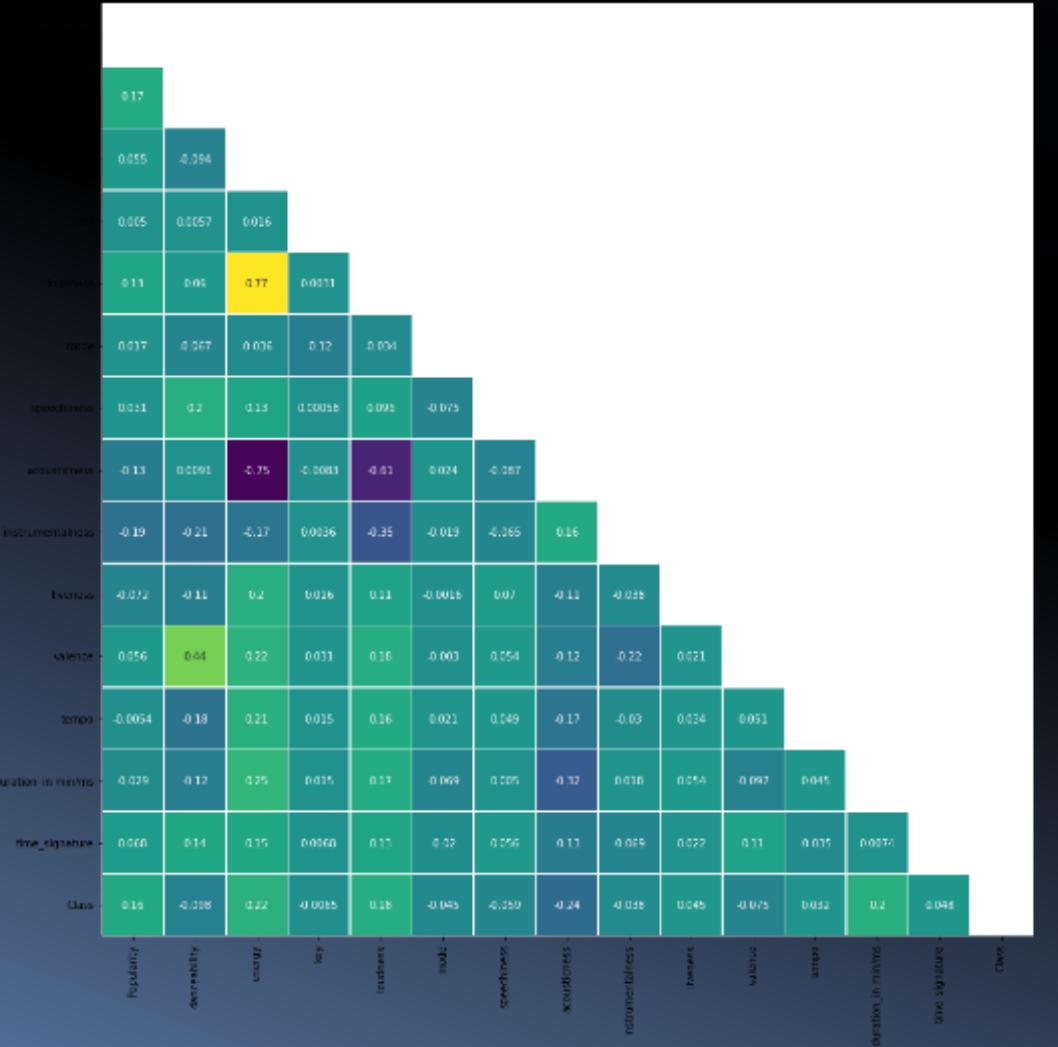
Exploración de Variables

Para esta sección se hace un análisis por variable y entre variables para poder encontrar datos atípicos y relaciones entre variables que puedan mejorar el desempeño de los modelos. En la tabla de correlaciones se puede apreciar que la mayoría de las variables no tiene una correlación lineal. A excepción de la variable de energía y de ruido(loudness) que tienen un coeficiente de Pearson de 0.77.



Análisis de Viabilidad

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos



Exploración de Variables

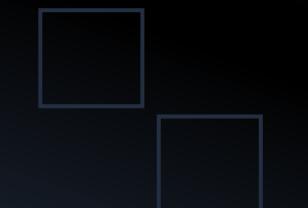
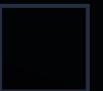
Una metaheurística inspirada en los principios de la selección natural

Análisis de Viabilidad

Implementación de los algoritmos genéticos

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Se decidió hacer una optimización usando algoritmos genéticos como método para optimizar el F1 Score general. De este modo se implementó los algoritmos genéticos usando la librería de Python TPOT, la cual permite desarrollar algoritmos genéticos sobre los algoritmos de machine learning en función a la métrica seleccionada.

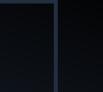
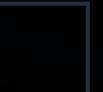
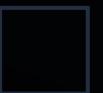


Análisis de Viabilidad

Implementación de los algoritmos genéticos

Función Objetivo

Para el diseño de algoritmos genéticos se planteo como función objetivo o de fitness la métrica de desempeño del F1 Score, los individuos de cada población serian los algoritmos de machine learning que son más populares hoy en día. Entre estos se puede encontrar: Random Forest, XGBoost, Maquinas de soporte vectorial, regresiones logísticas, etc..



Análisis de Viabilidad

Implementación de los algoritmos genéticos

Población Inicial

Para la selección de la población inicial se dejó que la librería seleccionara los algoritmos predefinidos para este, además para asegurar replicabilidad en los resultados se fijo la semilla de aleatoriedad que selecciona estos algoritmos.



Análisis de Viabilidad

Implementación de los algoritmos genéticos

Cruzamiento

Para el cruce de los individuos, se tiene predefinido que la mitad de los algoritmos se cruzaran un porcentaje de la población total que ya está predefinido, siempre y cuando los algoritmos sean de la misma naturaleza (es decir que sean los mismos algoritmos, pero con diferentes hiper parámetros).



Análisis de Viabilidad

Implementación de los algoritmos genéticos

Mutación y Selección...

Para la mutación de los algoritmos se toma un porcentaje predefinido, lo que cambia serán un hiperparámetro de los algoritmos que fueron seleccionados.

Después de haber realizado la mutación y el cruce de los Algoritmos se hará una selección por ruleta donde los algoritmos que tengan mejor función objetivo tendrán mayor probabilidad a ser seleccionados para siguiente generación. Este proceso se repite la cantidad de generaciones que se haya especificado.

Análisis de Viabilidad

Implementación de los algoritmos genéticos

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Como línea base el algoritmo genético predeterminado tendrá :

- Número de generaciones : 5
- Población inicial : 10
- Porcentaje de cruce y mutación : 90% y 10%

Análisis de Viabilidad

Resumen de los resultados

Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos

Escenario	Escenario	Modelo	Best F1 Score
1	Línea Base	Random Forrest (Predeterminado)	0.47
2	Número de generaciones:30	XGBoost (max_depth=4, learning rate=0.1, min_child_weight=3)	0.54
3	Número de individuos población:50	Random Forrest(características máximas = 025)	0.53
4	Porcentaje de cruce y mutación : 50% y 50%	Ensamble de algoritmos con XGBoost de diferentes características.	0.52
5	Mejor modelo redes neuronales	Redes neuronales con características del texto	0.57

Análisis de Viabilidad

Resumen de los resultados - Matriz de confusión del mejor algoritmo

	precision	recall	f1-score	support
Acoustic/Folk	0.65	0.55	0.60	532
	Alt	0.50	0.42	394
	Alt_Music	0.41	0.26	767
	Blues	0.92	0.93	172
	Country	0.47	0.71	1463
	HipHop	0.70	0.74	465
	Indie	0.44	0.02	402
	Instrumental	0.52	0.54	771
	Metal	0.62	0.62	116
	Pop	0.78	0.77	121
Rock	Rock	0.69	0.76	196
	accuracy		0.54	5399
	macro avg	0.61	0.57	5399
	weighted avg	0.54	0.54	5399

Análisis de Viabilidad

Conclusiones...

1

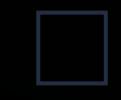
Se puede concluir que la calidad de los datos importa más que el algoritmo utilizado para la clasificación de las variables objetivo. Se puede apreciar que todos los algoritmos de machine learning ejecutados con diferentes configuraciones dan resultados similares a los algoritmos genéticos. Es decir, existe un límite en los resultados obtenido por la información procesada. En este caso, ningún algoritmo con las variables iniciales superó el 0.55, lo que sugiere que con las variables actuales este es el límite de efectividad.

Análisis de Viabilidad

Conclusiones...

2

Por el contrario de utilizar modelos de aprendizaje profundo que son denominados cajas negras, se pueden usar algoritmos con una interpretabilidad mayor y obtener resultados muy similares a los del aprendizaje profundo



3

En este caso, el mejor algoritmo encontrado fue ejecutado con 30 generaciones y 10 individuos cada una, lo cual puede significar que entre más generaciones los algoritmos encontrarán mejores resultados. Sin embargo, estos resultados están ligados a los datos de entrada y si se aplica el mismo algoritmo genético a una base de datos diferente estos resultados podrán cambiar significativamente.

**Sistema de Clasificación de Géneros Musicales
para plataformas de Streaming mediante
Algoritmos Genéticos**

□ □ □ **Gracias** □ □ □