



Anotado semiautomático de eventos de emergencia reportados en X basado en aprendizaje automático y computación evolutiva

Jesús Zambrano-Zambrano
Universidad Técnica de Manabí, UTM
jzambrano1217@gmail.com
Portoviejo, Ecuador.

Joel Garcia-Arteaga
Universidad Técnica de Manabí, UTM
jgarcia5169@utm.edu.ec
Portoviejo, Ecuador.

Jorge Parraga-Alava
Universidad Técnica de Manabí, UTM
jorge.parraga@utm.edu.ec
Portoviejo, Ecuador

Resumen

Los corpus de texto relacionados con incidencias ciudadanas son escasos y el proceso de anotación, para determinar si se trata de una emergencia, normalmente se realiza de forma manual a través de anotadores humanos. Esta forma de realizar el proceso a menudo produce resultados aceptables, pero son más lentos, costosos, y factibles únicamente para conjuntos con volumen pequeño o que no requieren procesamiento en tiempo real. Realizar anotaciones con anotadores humanos para corpus de emergencias, las que por su naturaleza requieren acciones inmediatas, no parece ser ideal, sobre todo cuando se reportan en redes sociales como X. En este artículo se propone un enfoque para el anotado semiautomático de eventos de emergencia reportados en X, basado en aprendizaje automático y computación evolutiva. Se utilizó la metodología CRISP-DM con seis etapas: comprensión del problema, recolección, y comprensión de datos, modelado, validación y despliegue. Los experimentos computacionales muestran un buen desempeño cuando se utiliza como modelo el clasificador de vectores de soporte lineal (LSVC), cuyos hiper-parámetros son optimizados con un algoritmo de computación evolutiva. Se alcanzaron valores promedio de 0.976 y 0.963 de F1-Score, así como 0.96 y 0.97 del coeficiente correlación de Matthews, para anotar la etiqueta de emergencia en +170 mil tweets para la clasificación binaria y multiclase, respectivamente. Se concluye que es posible tener un enfoque eficaz para asignar la etiqueta de emergencia a los textos publicados en la red social X, gracias a la optimización de los valores de los parámetros de los modelos de clasificación de aprendizaje automático.

Palabras clave: Anotado semiautomático, algoritmos genéticos, aprendizaje automático, X, corpus.

Semi-automatic annotation of emergency events reported on X based on machine learning and evolutionary computing

ABSTRACT

Text corpora related to citizen incidents are scarce and the annotation process, to determine whether they are emergencies, is usually done manually by human annotators. This annotation method yields acceptable results, but it is slower, more expensive, and only feasible for data sets that are small in volume or do not require real-time processing. Annotating emergency corpora with human annotators, which require immediate actions by their nature, does not seem to be ideal, especially when they are reported on social networks such as X. This article proposes an approach for the semi-automatic annotation of emergency events reported on X, based on machine learning and evolutionary computation. The CRISP-DM methodology was applied with six stages: problem understanding, data collection and understanding, modeling, validation and deployment. The computational experiments show a good performance when the linear support vector classifier (LSVC) is used as a model, whose hyper-parameters are optimized with an evolutionary computation algorithm. Average values of 0.976 and 0.963 for the F1-Score, as well as 0.96 and 0.97 for the Matthews correlation coefficient, were obtained to assign the emergency tag to +170 thousand tweets for the binary and multiclass classification, respectively. It is concluded that it is possible to have an efficient approach to assign the emergency label to the texts published on the social network X, thanks to the optimization of the values of the parameters of the machine learning classification models.

Keywords: Ecuador, hyper-parameter, genetic algorithm, machine learning, X, corpus.

1. Introducción

En el campo del procesamiento del lenguaje natural (NLP), un corpus es una gran colección de textos que se almacenan y procesan para describir el lenguaje natural. Generalmente, este corpus contiene información interpretativa adicional, denominada “anotación”, que se incorpora comúnmente de forma manual por anotadores humanos. Esta forma de incorporar la anotación a menudo arroja resultados confiables, pero es más lenta, costosa y factible únicamente para pequeños corpus. Los corpus relacionados con emergencia urbana son escasos, pero tienen la particularidad de contener gran cantidad de datos que describen

diferentes maneras de manifestar un sentimiento o reportar una emergencia en redes sociales.

Según (Luque et al., n.d.), en Ecuador el uso de X durante la pandemia del COVID-19 permitió difundir sucesos emergentes de toda índole. Esta situación ocasionó que no solo las instituciones del estado hicieran publicaciones de estos eventos, sino que una gran cantidad de ciudadanos también publicaran contenido relacionado a diversos tipos de emergencias. Esto provocó que organizaciones como la Policía Nacional del Ecuador, Cuerpo de Bomberos o Servicios de Asistencia Médica, todos ellos coordinados generalmente por el Servicio Integrado de Seguridad ECU 911, debieran prestar atención a la información difundida a través de estos medios tecnológicos, toda vez que en ocasiones esta contiene información errónea y/o contenido irrelevante, lo cual se traduce en datos de mala calidad. Por lo que es de vital importancia poder detectar el tipo de contenido o etiqueta que corresponde a cada posteo generado en redes sociales.

La literatura investigada muestra que ha habido varios autores que ya han realizado trabajos de parametrización de modelos de machine learning, tal como es el caso de (Yang, L., & Shami, A., 2020), que encontró que la parametrización correcta de los modelos de machine learning tiene un fuerte impacto en el desempeño de estos. Trabajos como (Spasis, I., 2020) muestran que el anotado de texto es uno de los principales cuellos de botella en el entrenamiento de modelos de inteligencia artificial, y esto se debe a que la investigación se ha centrado en el entrenamiento de modelos para clasificación, más no en el entrenamiento para etiquetado de corpus.

A partir de lo anterior, es claro que para optimizar la detección de emergencia en redes sociales es imperioso contar con un proceso de anotado eficaz, de bajo costo y sobre todo que permita la automatización de tal detección. Por esta razón, en este artículo se presenta un enfoque para el anotado semiautomático de eventos de emergencia reportados en la red social X, basado en técnicas de aprendizaje automático y computación evolutiva.

2. Metodología (Materiales y métodos)

2.1 Metodología CRISP-DM.

Para la realización de esta investigación se utilizó la metodología de Cross Industry Standard Process for Data Mining (CRISP-DM) (Wirth, 2000) debido a que proporciona características idóneas para un proyecto de análisis de datos. Las etapas de la metodología se detallan a continuación:

2.1.1 Entendimiento del problema.

El problema que se desea abordar es la falta de un enfoque capaz de asignar la etiqueta de emergencia de forma automatizada a gran cantidad de eventos reportados en tiempo real en todo el territorio ecuatoriano a través de la red X. Al lograr esto se dispondrá de un corpus de texto de incidentes que permita entrenar clasificadores para identificar nuevas emergencias producidas en tiempo real. Si bien actualmente existen muchos modelos de inteligencia artificial capaces de clasificar eventos como desastres naturales o emergencias en general, muchos de estos no son fáciles de parametrizar y requieren la intervención de un profesional capacitado para cumplir con esta tarea, además de que el proceso deberá ser realizado periódicamente para evitar problemas de sesgos en los datos, haciendo que tal proceso se vuelva inmanejable para un proceso manual humano. Por lo que también se requiere que los clasificadores pasen por un proceso automático de fijación de hiper-parámetros para obtener mejores resultados en la clasificación.

2.1.2 Recolección de los datos.

Para la realización de esta investigación se usó X como fuente de datos, que, de acuerdo a la literatura estudiada, es una de las redes sociales más utilizadas para proyectos de investigación y entrenamiento de modelos de aprendizaje automático e inteligencia artificial relacionado con análisis de texto.

La extracción de tweets se realizó en tiempo real usando un enfoque de captura desarrollado por los autores en una investigación previa (Garcia-Arteaga et al. 2022). En síntesis, el enfoque está basado en auto encoders para identificar palabras claves incluidas en tweets de emergencia, para después usar estas como filtros en la API de X, garantizando de este modo la obtención de un mayor porcentaje de datos relacionados con eventos de emergencia en español. En total durante dos meses (febrero-marzo 2022) se recolectaron cerca de 170 mil tweets sobre posibles eventos de emergencia en Ecuador, siendo las palabras más frecuentes “homicidio”, “asalto”, “emergencia”, “ECU911”, etc.

Tabla 1. Datos utilizados en el estudio

Total de tweets	Tipo	Categoría
1234	Emergencia	Seguridad Ciudadana Gestión Sanitaria Tránsito y Movilidad Gestión de Siniestros Gestión de Riesgos Servicio Militar Servicios Municipales
2405	No emergencia	No emergencia

En la Tabla 1 se observan los datos recolectados con las clases y categorías de emergencia consideradas en el estudio.

2.1.3 Comprensión de los datos.

En esta etapa se realizó el análisis exploratorio de datos (EDA por sus siglas en inglés *Exploratory Data Analysis*), el cual evidenció la existencia de un desbalance en la cantidad de datos para emergencia y no emergencia (Figura 1), teniendo una razón de 4 a 1, es decir, por cada tweet extraído hay cuatro que no corresponden a ninguna emergencia. Asimismo, los datos mostraron que existe

un desbalance considerable en cuanto a la distribución de los tipos de emergencia, siendo la categoría de *tránsito y movilidad* la que encabeza la lista con el 46 % de los tweets, seguida por *gestión de riesgos* con un 24 %, *seguridad ciudadana* con el 15 %, *servicios municipales* con el 12 %, y por último *gestión sanitaria* con el 3 % de los tweets, por lo que el modelo a utilizar debe tener en cuenta todos estos desbalances que pueden introducir sesgos en los datos.

También se realizó una etapa de preprocesamiento para convertir texto de lenguaje humano a un formato legible por computadora, para ello se eliminaron contenidos irrelevantes en el texto del tweet tales como URLs, signos de puntuación, etc., para luego realizar un proceso de tokenización.

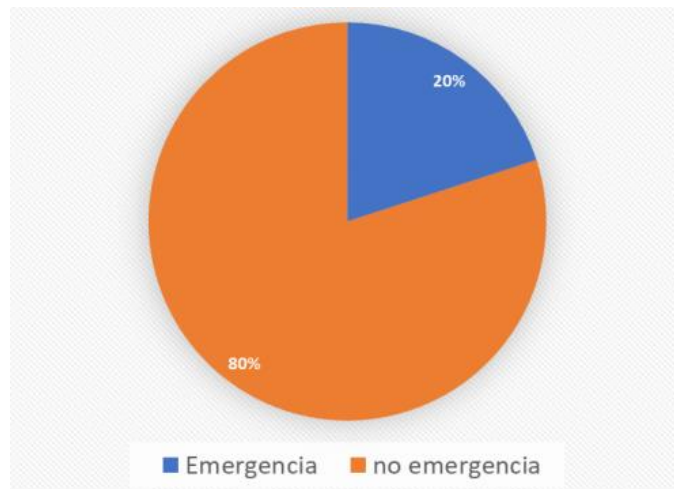


Figura 1. Proporción de datos de emergencia y no emergencia

2.1.4 Modelado.

El modelado realizado consistió en dos partes.

1. El uso de un clasificador que es un modelo de aprendizaje automático llamado clasificador de vectores de soporte lineal, LSVC por sus siglas en inglés (Mosquera et al., 2018). Se escogió este modelo dado que su aplicación en este contexto presenta resultados excelentes, en trabajos como (Rojo, 2022), donde se realizó la aplicación de un texto informal léxico

para el análisis de sentimientos, y se compararon los resultados y rendimiento de tres modelos distintos, siendo el LSVC el que obtuvo mejores resultados. En (Cànaves Alberti,2020), también LSVC se usó para la dotación automática de poemas, siendo este el que presentó el mejor rendimiento. En el trabajo de (Mares Giner, 2017), se utilizaron varios modelos tradicionales de machine learning para la clasificación de tweets, siendo el que destacó por encima de todos, el LSVC. También se selecciona LSVC porque se realizaron pruebas exhaustivas en otro trabajo de los autores (Joel Garcia-Arteaga, 2022).

2. El diseño de una arquitectura de algoritmo genético (modelo muy usado en computación evolutiva) capaz de parametrizar de forma eficiente el clasificador. Esta arquitectura cuenta con la estrategia de elitismo (Gallard Raúl, 2004), la cual asegura que los mejores individuos de todas las generaciones se preserven y sus genes puedan continuar mezclados y mutando para poder alcanzar menores tiempos de convergencia, asimismo la arquitectura prioriza la diversidad de los genes a través de una estrategia llamada crowding factor (Mengshoel et al., 2014), la cual consiste en agrupar individuos con una estructura genética similar en la generación actual, y decidir cuál de estos conservar mientras que los demás se descartan, aumentando la probabilidad de cruzar individuos de alto desempeño con individuos de bajo desempeño.

Teniendo en cuenta los factores previamente mencionados se ha creado una arquitectura de software que aproveche las características de los algoritmos genéticos y de la LSVC para de esta forma crear un modelo de clasificación que se mantiene aprendiendo a través del tiempo, considerando ciertos sesgos en los datos para poder brindar un desempeño óptimo a la hora de etiquetar texto. A continuación, se explica la arquitectura propuesta en la Figura 2.

El primer nodo cuenta con la ejecución del modelo Linear Support Vector Classification Genetic Algorithm (LSVCGA) descrito en el artículo (Joel Garcia-Arteaga, 2022) para la identificación de keywords para filtrar tuits de forma más

eficiente reduciendo la presencia de datos irrelevantes; el segundo nodo corresponde a una herramienta de extracción de tuits que se encargará de establecer una conexión con la API de X usando la salida del modelo LSVCGA; el tercer nodo es una herramienta para el entrenamiento periódico de la LSVC a través de la arquitectura de algoritmo genético propuesta para de esta forma reducir los sesgos en los datos a través del tiempo, como es el caso del data drift (Wood et al., n.d.) y por último se tiene el nodo de predicción, que es donde se aloja el modelo entrenado por el algoritmo genético y se encarga de clasificar tuits en emergencia y no emergencia, además de identificar la categoría a la que este pertenece (sanitaria, tránsito, siniestros, seguridad ciudadana, etc). Las flechas azules muestran el flujo de comunicación entre los diferentes nodos.

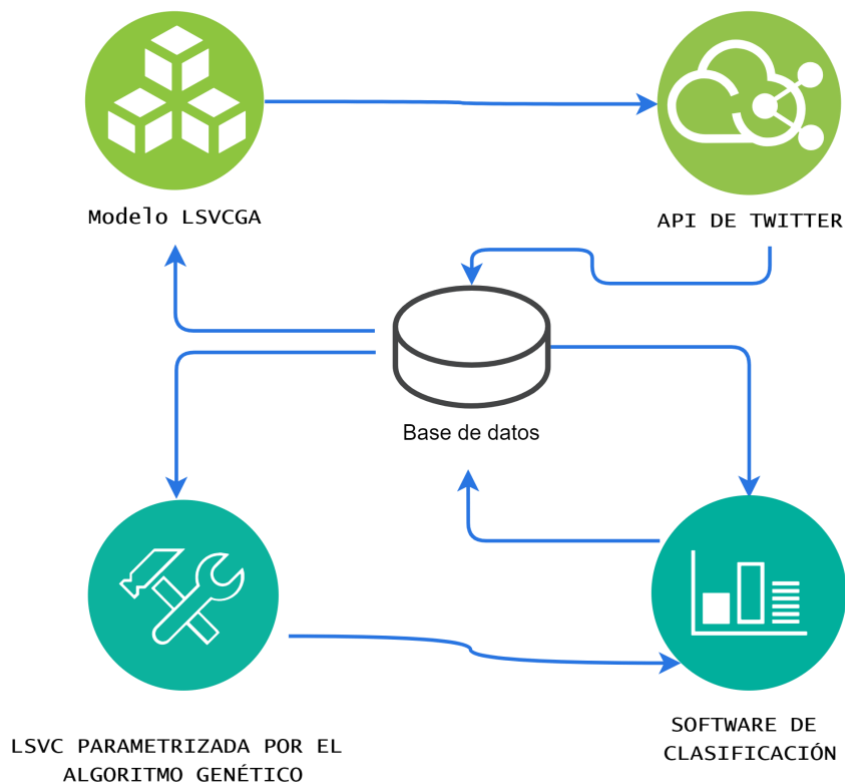


Figura 2. Arquitectura para la clasificación de textos de emergencias

2.1.4 Validación.

Para evaluar el rendimiento de la arquitectura del algoritmo genético al buscar el mejor conjunto de hiper-parámetros, se usa el coeficiente de correlación de Matthews (MCC) (Boughorbel et al., 2017). MCC puede interpretarse como un resumen de la matriz de confusión del proceso predictivo, siendo alto solo si el clasificador está funcionando bien tanto en los elementos negativos como en los positivos. Para evaluar el rendimiento del clasificador con sus valores por defecto, se utilizaron cuatro métricas de calidad comúnmente usadas en aprendizaje automático: accuracy, precisión, recall y F1-score.

2.1.5 Despliegue.

En el despliegue se creó un script simplificado en Python, como un sistema caja negra, que recibe en tiempo real los tweets extraídos con la API de X, y usando el clasificador optimizado con el algoritmo genético, se le asigna a cada tweet una etiqueta ("emergencia" y "no emergencia"). En el caso de los tweets de "emergencia", también se etiqueta la categoría de emergencia tal como se indica en la Tabla 1. El código fuente de esta etapa está disponible en <https://github.com/XxDextroxX/Tesis>

3. Resultados (análisis e interpretación de los resultados)

3.1. Análisis con respecto al clasificador

Los siguientes gráficos representan el rendimiento del LSVC, demostrando su desempeño con datasets de alta y baja calidad. Un dataset de alta calidad es aquel conjunto de datos idóneo para trabajar con modelos de machine learning, recordando que de la calidad de estos dependerá fuertemente los resultados que el modelo proporcione. En este contexto, el dataset de alta calidad contaba con un correcto preprocesamiento del texto, además de los métodos tradicionales como tokenizado, lematizado, eliminación de stopwords, etc., sumando que el proceso de etiquetado se realizó manualmente, optando por una especie de aprendizaje

por refuerzo, de modo que en las primeras fases de elaboración del dataset, se realizaban pruebas con nuevos conjuntos de datos sin etiquetar, y en los que el modelo fallaba, estos eran etiquetados manualmente, logrando así un dataset óptimo y de alta calidad para la problemática abordada. Por otro lado, se considera dataset de baja calidad a aquel conjunto de datos que no cumple con los estándares necesarios para poder entrenar un modelo de manera correcta, esto conlleva a obtener resultados no deseados. En este caso, vamos a considerar como dataset de baja calidad a un dataset donde cierto número de filas tienen una etiqueta incorrecta, recordando que en este trabajo también se implementó un modelo multiclase, no había suficientes ejemplares de calidad de cada etiqueta para el entrenamiento de modelo, sumando también que el preprocesamiento no abordaba de manera correcta los métodos tradicionales que se emplean al trabajar problemas de esta índole.

Tabla 2. Rendimiento LSVC dataset alta calidad.

	Accuracy	Precisión	Recall	F1-score
Binario	0.975	0.952	0.990	0.976
Multiclase	0.968	0.959	0.935	0.963

Fuente: Los autores (2022)

Tabla 3. Rendimiento LSVC + AG dataset alta calidad.

	Accuracy	Precisión	Recall	F1-score
Binario	0.973	0.957	0.995	0.967
Multiclase	0.971	0.951	0.958	0.962

Fuente: Los autores (2022)

Tabla 4. Rendimiento LSVC dataset baja calidad.

	Accuracy	Precisión	Recall	F1-score
Binario	0.829	0.863	0.790	0.825
Multiclase	0.912	0.854	0.896	0.90

Fuente: Los autores (2022)

Tabla 5. Rendimiento LSVC + AG dataset baja calidad.

	Accuracy	Precisión	Recall	F1-score
Binario	0.843	0.881	0.856	0.896
Multiclase	0.923	0.901	0.936	0.931

Fuente: Los autores (2022)

Como se puede observar, la calidad de los datos tiene un impacto bastante fuerte en qué tanta diferencia existirá entre utilizar algoritmos genéticos para la parametrización de la LSVC o utilizar sus parámetros por defecto, y es que si bien con el uso de datos de alta calidad (Tablas 2 y 3) no existe una diferencia significativa a priori entre ambos enfoques, usando datos de baja calidad (Tablas 4 y 5) sí que existe una mejora con el uso del algoritmo genético tanto para los escenarios de clasificación binaria como multiclase, siendo la métrica del recall una de las que más mejora obtuvo de la parametrización del algoritmo genético, alcanzando un score de 0.856 para clasificación binaria, lo cual representa una mejora considerable en la identificación de positivos verdaderos, asimismo en los datos de alta calidad se obtuvo un buen resultado en esta métrica, alcanzando el valor de 0.995, yendo en concordancia con el propósito de este trabajo que es ofrecer una herramienta para el etiquetado de datos de emergencia. La métrica de accuracy también obtuvo un buen resultado para los escenarios de clasificación binaria y multiclase con datos de baja calidad, alcanzando un promedio superior a 0.9 y en datasets de alta calidad llegando al valor de 0.97, aunque cabe mencionar

que esta métrica no obtuvo un mal rendimiento usando los parámetros por defecto que ofrece la LSVC.

Los algoritmos genéticos son una técnica de optimización inspirada en la evolución natural, que consiste en generar y seleccionar soluciones candidatas a un problema mediante operadores de cruce, mutación y selección, basados en la aptitud de cada solución (Holland, 1975). En este caso, se aplicaron algoritmos genéticos para encontrar el mejor conjunto de hiper-parámetros para el LSVC, que son valores que controlan el comportamiento y el rendimiento del modelo, tales como el tipo de kernel, el grado del polinomio, el coeficiente de regularización, etc. Al optimizar estos valores, se busca mejorar la capacidad de generalización y adaptación del modelo a los datos.

Una de las limitaciones que se encontró al aplicar este enfoque fue el tiempo de ejecución del algoritmo genético, que depende del tamaño de la población, el número de generaciones, la complejidad del modelo y la cantidad de datos. Para reducir este tiempo, se utilizaron estrategias de paralelización y distribución de la carga computacional, así como criterios de convergencia y elitismo. Otra limitación fue la calidad y cantidad de los datos disponibles, que influyen en la precisión y robustez del modelo. Para mejorar la calidad de los datos, se realizó un preprocesamiento adecuado del texto, eliminando ruido e información irrelevante, y se aplicó un proceso de anotación manual y semiautomática, usando un modelo previo para filtrar y etiquetar los datos. Para aumentar la cantidad de datos, se utilizó una herramienta de extracción de tweets en tiempo real, que se conecta con la API de X y usa palabras clave relevantes para la detección de emergencias.

Como posibles mejoras o extensiones de este trabajo, se podrían explorar otras técnicas de optimización, como la búsqueda en rejilla o la optimización bayesiana, para comparar su eficiencia y efectividad con los algoritmos genéticos. También se podrían probar otros modelos de clasificación, como las redes neuronales artificiales o las máquinas de vectores de soporte con kernel no lineal,

para evaluar su desempeño y capacidad de generalización con los datos de emergencia. Además, se podrían incorporar otras fuentes de datos, como otras redes sociales o medios de comunicación, para ampliar el corpus de texto y diversificar los tipos de emergencia. Finalmente, se podrían implementar sistemas de alerta temprana o de respuesta rápida, que usen el modelo de clasificación para identificar y atender las emergencias reportadas en las redes sociales.

3.2. Análisis con respecto al clasificador + algoritmo genético

El coeficiente de correlación de Matthews (MCC) es una medida de calidad que evalúa el rendimiento de un clasificador binario, teniendo en cuenta los cuatro valores de la matriz de confusión: verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. El MCC varía entre -1 y 1, donde 1 indica una predicción perfecta, 0 indica una predicción aleatoria y -1 indica una predicción inversa. El MCC es una buena métrica para evaluar el rendimiento del algoritmo genético y el LSVC, porque tiene en cuenta el balance de las clases y no se ve afectado por el cambio de escala de los datos.

En la Figura 3 se puede observar el rendimiento y velocidad de convergencia promedio y de elitismo de la arquitectura de algoritmo genético usada para parametrizar la LSVC, pudiéndose observar una convergencia con un puntaje de 0.96 en MCC pasada la generación 100 para los escenarios de clasificación binaria y multi clase, mostrándose en el eje horizontal el número de generaciones y en el eje vertical el MCC. La línea azul representa el rendimiento de la arquitectura mediante el promedio de los individuos que se encuentran en el salón de la fama (elitismo), mientras que la línea naranja representa el rendimiento promedio de todos los individuos a través de las generaciones.

Estos resultados indican que el algoritmo genético logra encontrar una solución óptima para el LSVC en un tiempo razonable, y que el elitismo ayuda a preservar y mejorar la calidad de los individuos. Además, se observa que el

rendimiento promedio y el de elitismo son estables y no presentan fluctuaciones significativas, lo que sugiere que el algoritmo genético no cae en óptimos locales ni sufre de sobreajuste.

Para comparar los resultados de la Figura 3 con los de otras técnicas de optimización o parametrización, se podrían usar las mismas métricas de calidad y los mismos datos, y aplicar métodos como la búsqueda en rejilla, la búsqueda aleatoria o la optimización bayesiana, que son algunos de los más usados en la literatura. Estos métodos consisten en explorar el espacio de hiper-parámetros de forma sistemática, aleatoria o probabilística, respectivamente, y seleccionar el mejor conjunto de valores que maximice una función objetivo, como el MCC o el F1-score. Algunas ventajas de los algoritmos genéticos sobre estos métodos son que pueden escapar de óptimos locales, explorar y explotar el espacio de búsqueda de forma eficiente, y adaptarse a cambios dinámicos en los datos. Algunas desventajas son que pueden requerir más tiempo de ejecución, más recursos computacionales y más ajustes de sus propios parámetros, como el tamaño de la población, la tasa de cruce, la tasa de mutación, etc.

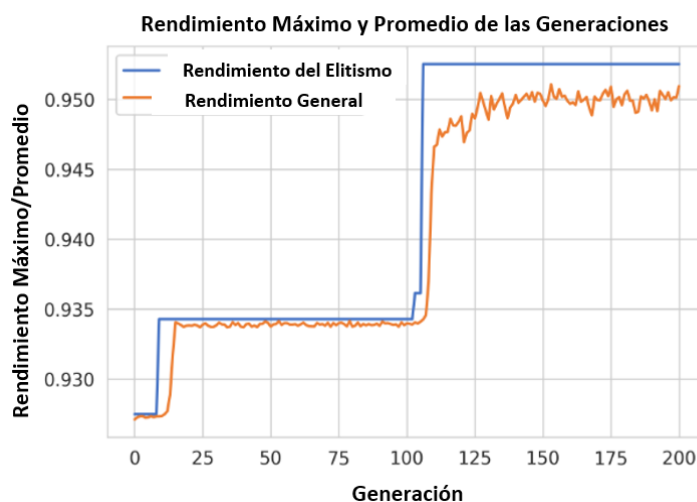


Figura 3: Rendimiento del algoritmo genético, para dataset binario.

Fuente: Los autores (2022).

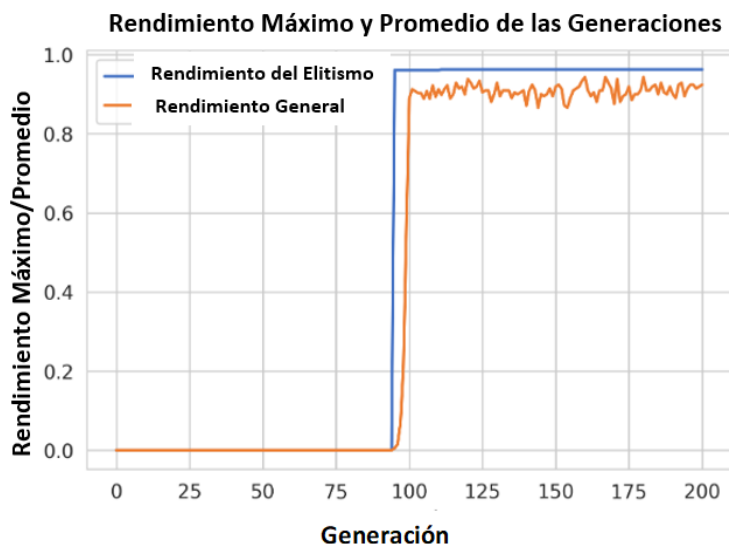


Figura 4: Rendimiento del algoritmo genético, para dataset multiclase.

Fuente: Los autores (2022).

Como se observa la parametrización resulta útil en este escenario de anotado semiautomático para datos de emergencia, pudiendo lograr etiquetar correctamente 8.5 tuits de cada 10 que entran con datasets de baja calidad, y etiquetando 9.9 tuits de cada 10 que entran si se usa un dataset de alta calidad, asimismo el algoritmo genético logra converger bastante rápido, por lo que la implementación de un flujo de clasificación usando AG + LSVC puede resultar bastante eficaz y eficiente.

4. Conclusiones

En este artículo se presentó un enfoque de anotado semiautomático de eventos de emergencia reportados en X basado en aprendizaje automático y computación evolutiva.

El uso de clasificadores estándar como el LSVC demuestran que pueden ser capaces de obtener buenos resultados etiquetado de textos, no obstante, es de suma importancia disponer de un conjunto de datos de buena calidad, puesto que, al no contar con uno, los resultados obtenidos no serán los deseados.

El uso del algoritmo genético permitió establecer los valores idóneos del clasificador, permitiendo obtener mejores resultados en cuanto a la determinación de la clase del tweet, así como de la categoría de la emergencia. Los resultados evidencian que no se necesita una cantidad elevada de iteraciones para que el modelo de algoritmo genético converja, pudiendo de esta manera implementar software que haga un uso eficiente de los recursos al no ejecutar iteraciones innecesarias.

Como trabajos futuros se puede probar la eficacia de otros modelos de la familia del cómputo evolutivo, así como la parametrización de más modelos de machine learning que pudiesen tener un desempeño superior al de los clasificadores de soporte lineal, teniendo en cuenta parámetros temporales como la época del año en la que se encuentra o la calidad del tuit. También se debe probar la eficiencia de los algoritmos genéticos parametrizando modelos más complejos como las Máquinas de Boltzmann o Redes neuronales profundas, que podrían dar resultados mucho más precisos a la hora de realizar clasificaciones.

5. Referencias

Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE, 12(6). doi: 10.1371/JOURNAL.PONE.0177678

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1). doi: 10.1186/S12864-019-6413-7

Ding, J., Data, X. L.-2018 I. I. C. on B., & 2018, undefined. (n.d.). An approach for validating quality of datasets for machine learning. leeexplore.lee.org. Retrieved from

García-Arteaga, J., Zambrano-Zambrano, J., Parraga-Alava, J., An effective approach for identifying keywords as high-quality filters to get emergency-implicated X Spanish data [Manuscript submitted for publication]. Computer Speech & Language.

Han, J. H., Choi, D. J., Park, S. U., & Hong, S. K. (2020). Hyperparameter Optimization Using a Genetic Algorithm Considering Verification Time in a Convolutional Neural Network. Journal of Electrical Engineering and Technology, 15(2), 721–726. doi: 10.1007/S42835-020-00343-7

Luque, A., Maniglio, F., Casado, F., & García-Guerrero, J. (n.d.). Transmedia Context and X As Conditioning the Ecuadorian Government's Action. The Case of the "Guayaquil Emergency" During the COVID-19 Pandemic. Raco.Cat, 2, 47–68.

Martínez-Rojas, M., ... M. del C. P.-F.-I. J. of, & 2018, undefined. (n.d.). X as a tool for the management and analysis of emergency situations: A systematic literature review. Elsevier. Retrieved from

Ranjit, M., Ganapathy, G., ... K. S.-2019 I. 12th, & 2019, undefined. (n.d.). Efficient deep learning hyperparameter tuning using cloud infrastructure: Intelligent distributed hyperparameter tuning with bayesian optimization in the cloud. leeexplore.lee.org. Retrieved from

Wood, J., Griffis, T., Meteorology, J. B.-A. and F., & 2015, undefined. (n.d.). Detecting drift bias and exposure errors in solar and photosynthetically active radiation data. Elsevier. Retrieved from

Rojo, V., Pollo-Cattaneo, M. F., & Britos, P. (n.d.). Slanglex-ar: aplicación de un léxico de lenguaje informal de Argentina para el análisis de sentimientos en español en X. *Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas y Accesibilidad*, 205.

Cànaves Alberti, M. (2020). *Datación automática de poemas mediante técnicas de aprendizaje automático*.

Mares Giner, J. M. (2017). *Visualización y seguimiento de acontecimientos en X* (Doctoral dissertation, Universitat Politècnica de València).

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316.

Spasic, I., & Nenadic, G. (2020). Clinical text data in machine learning: systematic review. *JMIR medical informatics*, 8(3), e17984.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).