Medición del desempeño de clasificadores usando atributos sintéticos polinomiales y selección de atributos con MrMR

L. A. Banda Peláez[1], K. M. Ramírez Vergara[2], A. López Chau[3], V. Trujillo Mora[4], C. O. González Moran[5]

[1]UAEMex; lbandap001@alumno.uaemex.mx,[2]UAEMex; kramirezv003@alumno.uaemex.mx, [3]UAEMex; alchau@uaemex.mx, [4]UAEMex; vtrujillom@uaemex.mx, [5]UAEMex; cogonzalezm@uaemex.mx.

***Área de participación**: Sistemas Computacionales*

Resumen

Las diferentes complejidades presentes en los datos perjudican el desempeño de los modelos predictivos. Entre las complejidades más comunes en los datos se encuentra el desbalance de clases, la presencia de casos atípicos traslape de clases, y alta dimensionalidad. Una de las formas de tratar con este problema es la creación y adición de atributos sintéticos a los datos, con la finalidad de mejorar el desempeño. En este artículo, se realiza una comparativa del comportamiento (en términos de la métrica F1-score) de seis clasificadores cuando se agregan atributos sintéticos de tipo polinomial. El objetivo de los experimentos es verificar si la creación de atributos sintéticos ayuda a lograr un mejor desempeño en comparación con los atributos originales.

***Palabras clave**: Clasificador, Métricas de desempeño, Atributos sintéticos, MrMR.*

Abstract

The different complexities present in the data impair the performance of predictive models. Among the most common complexities in data are class imbalance, the presence of outliers, class overlap, and high dimensionality. One of the ways to deal with this problem is to create and add synthetic attributes to the data, in order to improve performance. In this article, a comparison is made of the behavior (in terms of the F1-score metric) of six classifiers when polynomial-type synthetic attributes are added. The objective of the experiments is to verify if the creation of synthetic attributes helps to achieve better performance compared to the original attributes.

***Key words**: Classifier, Performance metrics, Synthetic attributes, MrMR.*

Introducción

Una de las tareas más importantes en minería de datos es la clasificación, la cual pertenece al tipo de aprendizaje supervisado. Existen varios métodos de clasificación, entre ellos se encuentran los árboles de decisión, máquinas de vectores soporte, redes neuronales, K-vecinos más cercanos y clasificador Bayesiano ingenuo. Cuando se aplican estos métodos de clasificación a un mismo conjunto de datos etiquetado, generalmente se obtienen resultados de desempeño diferente, es decir, para cada método de clasificación empleado sobre un mismo conjunto de datos, se presentan resultados diferentes dependiendo de las complejidades del modelo predictivo aplicado.

Entre los principales tipos complejidades que se pueden encontrar en datos etiquetados se pueden nombrar los siguientes [1]:

a) *Elevado número de instancias o muestras en el conjunto de datos*. Provoca que no se pueda cargar en memoria principal todo el conjunto de datos.
b) *Número muy reducido de muestras*. No existen suficientes muestras para generar modelos predictivos con buen desempeño.

c) *Alta dimensionalidad.* Cuando se tiene un elevado número de atributos, es posible que varios de ellos perjudiquen el desempeño de los modelos predictivos, debido a que son irrelevantes o redundantes.
d) *Traslape de clases*. El traslape de clases es un problema complejo, provoca que los clasificadores obtengan una tasa de error de predicción alta.
e) *Desbalance.* Sucede cuando hay un gran número de muestras de una clase, y una proporción significativamente más pequeña de otra clase.
f) *Casos anómalos*. Son muestras muy diferentes al resto, de acuerdo con algún criterio estadístico, esto puede hacer que los modelos predictivos se sesguen.

Estas complejidades dañan la exactitud de las predicciones de los clasificadores, por lo que se han diseñado estrategias orientadas a mejorar el desempeño de los métodos de clasificación, por ejemplo, la creación y adición de atributos sintéticos a los datos. Las máquinas de vectores de soporte (SVM) son uno de los algoritmos de clasificación y regresión más potentes y robustos en múltiples campos de aplicación; generan indirectamente una elevada cantidad de atributos, a través de funciones no lineales (llamadas funciones Kernel), para obtener un hiperplano de clasificación en un espacio de características de dimensión potencialmente infinita [2]. Otros métodos, como las redes neuronales totalmente conectadas, crean nuevos atributos a través de capas ocultas, mediante las salidas de funciones de activación que toman como argumentos los productos internos entre salidas de capas anteriores y los pesos sinápticos de la capa actual.

Las características sintéticas se generan a partir de los atributos existentes de manera implícita o explícita. Los métodos implícitos de generación de atributos sintéticos son efectivos para reducir los efectos de la complejidad de los datos sobre el desempeño de los clasificadores, aunque en otro sentido, tienen un funcionamiento menos comprensible para los humanos. La generación explícita de atributos sintéticos se realiza mediante operaciones entre atributos, como las sumas, productos, potencias y algoritmos. Este tipo de métodos considera la importancia de los atributos y descarta los atributos irrelevantes, con el objetivo de conocer explícitamente las transformaciones realizadas a los datos, sin embargo, el costo computacional puede llegar a ser elevado [3].

En este artículo se realiza un estudio exploratorio sobre el efecto que tiene la generación y adición de atributos sintéticos de tipo polinomial sobre el desempeño de seis métodos de clasificación muy conocidos. Para validar los resultados, se usan seis conjuntos de datos disponibles públicamente en internet. La metodología propuesta en el artículo consta de tres secciones. La primera de ellas describe las métricas más populares para evaluar el desempeño de clasificadores; la segunda sección menciona la metodología empleada y aplicada en esta investigación. Finalmente, se muestran los resultados tres experimentos aplicados a seis conjuntos de datos, con la finalidad de conocer si el uso se atributos sintéticos favorecen o no a los modelos predictivos.

Antecedentes

Métricas de desempeño de clasificadores

La tarea de clasificación en aprendizaje automático es ampliamente usada en problemas de diversas áreas, para poder aplicar un clasificador en ambientes reales, es necesario medir su desempeño en ambientes realistas. La técnica empleada comúnmente para validar los resultados consiste en usar dos tipos de muestras. El primer conjunto de datos etiquetados (*datos de entrenamiento)*, es para generar un modelo predictivo (el clasificador propiamente), el segundo conjunto *(datos de prueba)*, sirve para evaluar las predicciones realizadas por el clasificador; usualmente se utiliza un 80% de los datos para entrenamiento, y el 20% de los datos para pruebas.

Para medir la calidad de las predicciones de un clasificador, existe una variedad de métricas, la mayoría de ellas se obtienen de las estadísticas condensadas en una matriz confusión, cuya estructura general se muestra en la Figura 1 [4].

	Predicción positiva	**Predicción negativa**
Clase positiva	Verdaderos positivos (VP)	Falsos negativos (FN)
Clase negativa	Falsos positivos (FP)	Verdaderos negativos (VN)

Figura 1. Matriz de confusión para clasificación binaria.

Las matrices de confusión se pueden extender al número de clases que contenga el conjunto de datos. Con base en ella, se definen métricas de desempeño de un modelo predictivo como la exactitud, la precisión, la exhaustividad y F1-score.

La exactitud se calcula dividiendo la cantidad de predicciones correctas por el total de los datos [5]. Así mismo, la tasa de error mide los errores que el clasificador comete [4]. La forma de calcular estas métricas se muestra en la ecuación (1):

$$Exactitud = \frac{VP\#VN}{VP\#FN\#FP\#VN} \qquad Tasa\ de\ error = \frac{FP\#FN}{VP\#FN\#FP\#VN} \quad (1)$$

La precisión (ecuación 2) es una métrica caracterizada por utilizar las predicciones positivas realizadas, está dada por un cociente que toma en cuenta el número de predicciones positivas y el total de todas las predicciones positivas [4].

$$Precisión = \frac{Verdaderos\ positivos}{Verdaderos\ positivos\#Falsos\ positivos} \quad (2)$$

La métrica de exhaustividad tiene como objetivo principal, minimizar los falsos negativos, es decir, obtener un porcentaje de las predicciones positivas realizadas en relación con el total de predicciones positivas que posiblemente se realizaron [4], la ecuación (3) muestra la forma de calcular la exhaustividad.

$$Exhaustividad = \frac{VerdaderosPositivos}{VerdaderosPositivos\#Falsos\ \ Ne2ativos} \quad (3)$$

Para el cálculo de F-measure (o F1-score) se usa una media armónica, donde el numerador es el producto de la precisión y la exhaustividad multiplicada por dos, mientras que le denominador consta de la suma entre la precisión y la exhaustividad (ver ecuación (4)). Al usar los porcentajes de la precisión y la exhaustividad permite simplificar en un solo valor el rendimiento del clasificador [6].

$$F - measure = \frac{3\times Precisión\times Ex:austividad}{Precisión\#Ex:austividad} \quad (4)$$

Generación de atributos sintéticos polinomiales

Los atributos sintéticos son la creación de nuevas características artificiales, basándose en los datos reales. El propósito principal de agregar este tipo de atributos a los datos es mejorar la calidad del modelo de clasificación [7]. Existen diversas maneras de generar atributos sintéticos a partir de los atributos presentes en un conjunto de datos. En este trabajo, inspirados en el kernel polinomial de las SVM (ver ecuación (5)), se eligió la utilización de formas polinómicas.

$$K(x, y) = (x^T y + c)^d \quad (5)$$

Los atributos sintéticos que se generan tienen la forma mostrada en la ecuación (6):

$$a^{<} = \prod_{i}^{i\in C} x_i \quad (6)$$

Donde

a^j: es el atributo sintético j

$x^{\$}$: es el atributo i-ésimo del conjunto de datos

C: es una de las $\frac{(n^{\blacktriangledown}\bigcirc k)!}{(n^{\blacktriangledown}\bigcirc)!k!}$ posibles combinaciones (sin reemplazo), siendo n los mejores atributos y k el grado del polinomio.

Por ejemplo, si un conjunto de datos tiene tres atributos ($x_{(}$, x_2, x_3), y $k = 2$, entonces los atributos sintéticos generados serían $x_{(}x_{(}$, x_2x_2, x_3x_3, $x_{(}x_2$, $x_{(}x_3$, x_2x_3. Es importante aclarar, que solo se consideran atributos de tipo numérico entero o real. Debido a que la cantidad de atributos generados mediante la ecuación (6) puede llegar a ser muy grande, se aplica una técnica de selección de atributos.

Selección de atributos con MrMR

La selección de atributos con MrMR (Mínima redundancia – Máxima Relevancia) es un método reciente, que ha demostrado su efectividad. Este método permite seleccionar las características de un conjunto de datos que tienen mayor poder para discriminar entre clases, y simultáneamente, reduce la redundancia de los atributos [8]. En cada iteración MrMR evalúa el atributo f mediante el cociente mostrado en la ecuación (7).

$$score_{\mathrm{i}}(f) = \frac{relevance(@\ |tar2et)}{redundancy(@\ |\ @eatures\ selected\ until\ \mathrm{i–E})} \quad (6)$$

La relevancia del atributo f se calcula mediante la dependencia estadística que cada atributo tiene con la clase [9]. La redundancia es con respecto a las características ya elegidas hasta el paso i-ésimo de MrMR.

Metodología

Conjunto de datos

La Tabla 1 muestra un resumen de los conjuntos de datos usados en los experimentos. Todos estos datos se encuentran disponibles públicamente en internet [10].

Tabla 1. Resumen de los conjuntos de datos usados en los experimentos.

Conjunto de datos	Total de muestras	Total de atributos	Clases
cmc	1473	9	3
dermatology	366	34	6
ecoli	336	8	8
glass	214	9	7
haberman	306	3	2
raisin shuttle	14500	9	7

La Figura 1 muestra un resumen gráfico de la metodología empleada en este artículo. La estructura principal consiste en cuatro bloques, explicados a continuación.

El primer paso denominado ‘preparación’, corresponde a la lectura de los datos. En esta etapa se separan los atributos de las clases, además se implementa un filtro que tiene como objetivo seleccionar los atributos de tipo numérico (enteros o reales), y eliminar a los que no son numéricos, como nominales, categóricos u ordinales.

Experimento I: Desempeño de clasificadores con todos los atributos

En este paso, se separa pseudo-aleatoriamente el conjunto de datos, analizado en dos subconjuntos disjuntos: el primer subconjunto corresponde al de entrenamiento (80% de los datos) y el segundo subconjunto al de prueba (20% restante). El conjunto de datos de entrenamiento (X_{train}, Y_{train}) es usado para generar un modelo predictivo o clasificador, mientras que el de prueba (X_{test}, Y_{test}) es empleado para calcular la precisión, exhaustividad y F1-score del modelo. La finalidad de calcular el desempeño de cada clasificador usando todos los atributos numéricos es para compararlo con el conjunto de datos con los atributos sintéticos añadidos. Es importante mencionar que los desempeños se reportan considerando 30 repeticiones de cada experimento, usando los mismos datos en todos ellos, pero seleccionado muestras diferentes en cada repetición.

Experimento II: Desempeño de clasificadores con atributos seleccionados con MrMR

En el segundo experimento se seleccionan los k mejores atributos del conjunto de datos de entrenamiento con MrMR. Esos mismos atributos son elegidos del conjunto de datos de prueba. Posteriormente, se genera un modelo predictivo y se evalúa su desempeño. El propósito de este experimento es inspeccionar el efecto de MrMR sobre el desempeño del clasificador.

Experimento III: Desempeño de clasificadores con atributos sintéticos añadidos

En este experimento, se generan atributos sintéticos a partir de los mejores atributos seleccionados en el experimento II. Como se mencionó anteriormente, los atributos sintéticos generados son polinomiales, es decir, productos de variables numéricas presentes en el conjunto de datos original. En este experimento la carga computacional es elevada, debido a que la cantidad de atributos generados es muy grande. Para evitar el consumo excesivo de memoria, se decidió aplicar MrMR a los atributos sintéticos. De esta manera, se generan modelos predictivos usando los atributos originales y los sintéticos más relevantes.

Almacenamiento de resultados

En la fase de almacenamiento se reestructuran todos los datos obtenidos por cada prueba. Los resultados reportados son la media y desviación estándar de la precisión, exhaustividad y F1-score de los 30 experimentos.

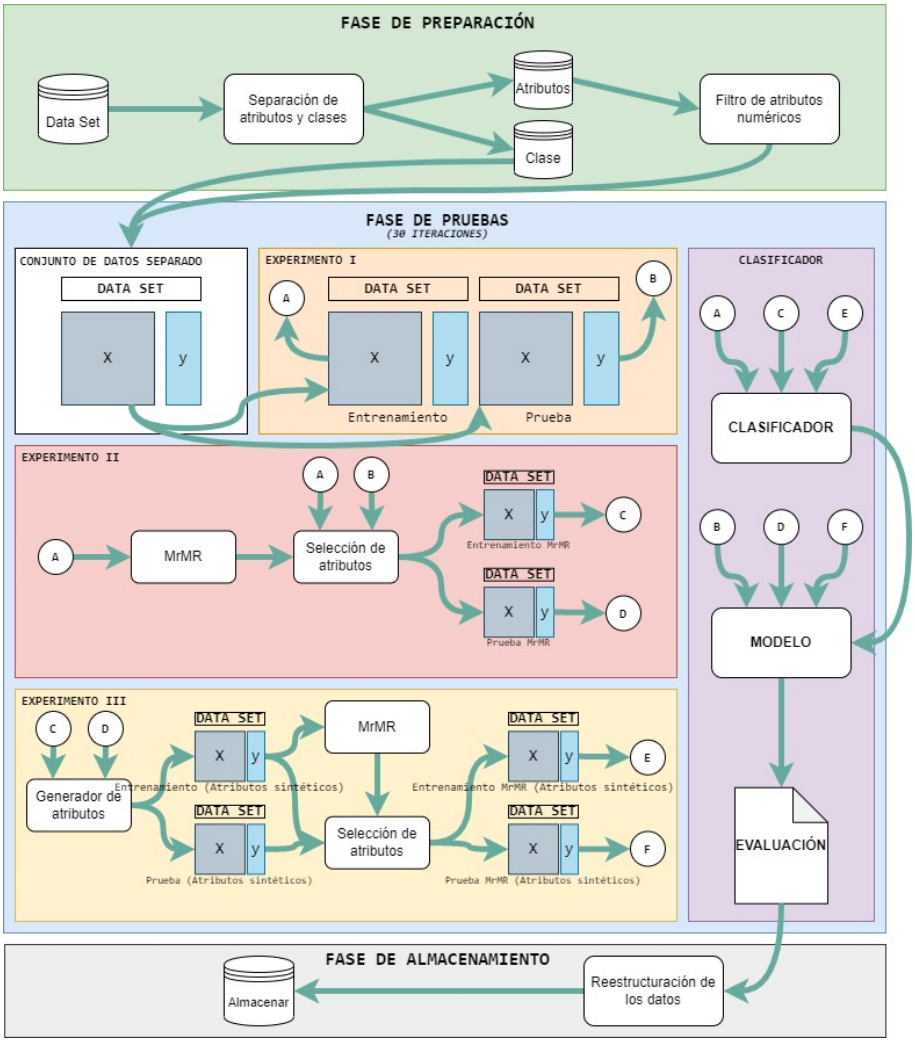

Figura 3. Diagrama representativo del funcionamiento del algoritmo de un modelo de clasificación. Fuente: Elaboración propia.

Resultados

Se midieron los desempeños de los clasificadores K-NN, regresión logística, árbol de decisión, SVM, perceptrón multicapa (MLPC) y Naive Bayes con cada uno de los conjuntos de datos, para cada experimento. Debido a la gran cantidad de datos generados, se decidió mostrar solamente el valor de F1-score. Para cada experimento, se ajustaron los principales parámetros de cada clasificador aplicando el método de búsqueda exhaustiva por rejilla. Todos los experimentos fueron ejecutados en una computadora.

Resultados Experimento I

La tabla 2 muestra los valores de F1-score, precisión y exhaustividad para cada conjunto de datos.

Tabla 2. Desempeños de clasificadores (F1-score/STD) con todos los atributos

Conjunto de datos	KNN	SVM	Logistic Regression	Decision Tree	Naive Bayes	Neural Network
cmc	0.47/0.02	0.50/0.02	0.51/0.02	0.49/0.01	0.48/0.02	0.50/0.03
dermatology	0.97/0.02	0.98/0.02	0.98/0.02	0.96/0.03	0.88/0.02	0.94/0.02
ecoli	0.70/0.03	0.88/0.02	0.86/0.02	0.79/0.04	0.80/0.04	0.70/0.03
glass	0.60/0.03	0.58/0.07	0.51/0.06	0.64/0.03	0.26/0.10	0.56/0.06
haberman	0.67/0.02	0.66/0.03	0.67/0.02	0.70/0.04	0.72/0.05	0.66/0.04
raisin shuttle	1.00/0.00	0.99/0.00	0.95/0.00	1.00/0.00	0.92/0.01	0.99/0.01

En los conjuntos de datos cmc, glass y ecoli es donde se observa que los clasificadores alcanzan un bajo desempeño. Por otra parte, el conjunto de datos raisin_shuttle es en el cual los clasificadores presentan un mayor valor de F1-score.

Resultados Experimento II

Las tablas 3a presenta los desempeños de los métodos de clasificación analizados. Se observa que, para 15 casos, se mejoró el desempeño ligeramente.

Tabla 3a. Desempeños de clasificadores (F1-score/STD) con selección de atributos

Conjunto de datos	KNN	SVM	Logistic Regression	Decision Tree	Naive Bayes	Neural Network
cmc	0.46/0.02	0.52/0.03	0.52/0.02	0.43/0.02	0.51/0.02	0.52/0.03
dermatology	0.98/0.01	0.98/0.02	0.98/0.01	0.95/0.01	0.82/0.05	0.95/0.04
ecoli	0.86/0.03	0.87/0.03	0.88/0.03	0.80/0.04	0.76/0.07	0.71/0.05
glass	0.67/0.06	0.60/0.02	0.58/0.03	0.63/0.05	0.47/0.11	0.69/0.06
haberman	0.68/0.04	0.70/0.04	0.68/0.05	0.74/0.04	0.72/0.03	0.65/0.02
raisin shuttle	1.00/0.00	0.98/0.00	0.93/0.00	1.00/0.00	0.94/0.00	0.99/0.01

La tabla 3b muestra los mejores parámetros encontrados para cada método de clasificación. En la primera columna se muestra la cantidad de atributos seleccionados en cada conjunto de datos.

Tabla 3b. Parámetros de los clasificadores con selección de atributos

Conjunto de datos Atributos Seleccionados	KNN	SVM	Logistic Regression	Decisión Tree	Neural Network
cmc MrMR:9	K: 13	C: 5, degree:3, kernel: poly	C: 10	criterion: entropy	activation: relu
dermatology MrMR: 24	K: 13	C: 1, degree: 4, kernel: poly	C: 1	criterion: entropy	activation: relu
Ecoli MrMR: 5	K: 7	C: 10, gamma: auto, kernel: sigmoid	C:10	criterion: gini	activation: tanh
glass MrMR: 5	K: 3	C: 10, degree: 2, kernel: poly	C: 10	criterion: entropy	activation: identity
haberman MrMR: 1	K: 11	C: 10, gamma: auto, kernel: sigmoid	C: 0.01	criterion: gini	activation: logistic

raisin shuttle MrMR: 3	K: 3	C: 100, degree: 4, kernel: poly	C: 10	criterion: gini	activation: relu
cmc MrMR:9	K: 13	C: 5, degree:3, kernel: poly	C: 10	criterion: entropy	activation: relu

Resultados Experimento III

Tabla 4a. Desempeños de clasificadores (F1-score/STD) con selección y generación de atributos sintéticos polinomiales

Conjunto de datos	KNN	SVM	Logistic Regression	Decision Tree	Naive Bayes	Neural Network
cmc	0.48/0.02	0.52/0.03	0.54/0.03	0.44/0.04	0.49/0.03	0.54/0.02
dermatology	0.94/0.03	0.95/0.03	0.94/0.03	0.91/0.02	0.83/0.03	0.95/0.04
ecoli	0.73/0.03	0.71/0.04	0.69/0.04	0.71/0.02	0.76/0.02	0.71/0.04
glass	0.94/0.02	0.87/0.03	0.75/0.03	0.98/0.02	0.96/0.01	0.73/0.03
haberman	0.73/0.05	0.67/0.03	0.66/0.03	0.64/0.03	0.71/0.04	0.67/0.03
raisin shuttle	1.00/0.00	0.93/0.00	0.90/0.00	1.00/0.00	0.94/0.00	0.99/0.01

La tabla 3b muestra los mejores parámetros encontrados para cada método de clasificación. En la primera columna se muestra la cantidad de atributos seleccionados en cada conjunto de datos, y el grado polinomial de los atributos sintéticos.

Tabla 4b. Parámetros de los clasificadores con selección y generación de atributos sintéticos polinomiales.

Conjunto de datos Atributos Seleccionados	KNN	SVM	Logistic Regression	Decisión Tree	Neural Network
cmc MrMR: 9 Polinomio: 6	K: 17	C': 100, 'gamma': 0.01, 'kernel': 'sigmoid	C: 1	criterion: entropy	activation: identity
dermatology MrMR: 20 Polinomio: 1	K: 7	C: 100, gamma: 0.01, kernel: sigmoid	C: 1	criterion: gini	activation: relu
Ecoli MrMR: 8 Polinomio: 2	K:11	C: 10, gamma: 1, kernel: sigmoid	C: 10	criterion: entropy	activation: identity
glass MrMR: 2 Polinomio: 2	K: 3	C: 100, gamma: 0.1, kernel: sigmoid	C: 10	criterion: entropy	activation: identity
haberman MrMR: 2 Polinomio: 2	K: 7	C: 100, gamma: 0.1, kernel: sigmoid	C: 10	criterion: gini	activation: logistic
raisin shuttle MrMR: 2 Polinomio: 2	K: 3	C: 100, degree: 5, kernel: poly	C: 10	criterion: gini	activation: relu

Discusión

Con base en los resultados obtenidos, se observa lo siguiente:

- La metodología propuesta es correcta y adecuada para evaluar el efecto de nuevos atributos agregados a los datos, ya que permite evaluar de manera objetiva si estos atributos logran contribuir a enfrentar las complejidades en los datos, mediante la medición del desempeño de los métodos de clasificación.
- La adición de atributos sintéticos polinomiales tiene un efecto positivo (aunque pequeño) sobre el desempeño de los clasificadores con los conjuntos de datos analizados.
- La técnica de selección de atributos por sí misma, también tuvo un efecto positivo y pequeño, ya que los valores de F1-score mejoraron un poco en algunos casos, pero no fue así en la mayoría.

Por lo tanto, se sugiere lo siguiente:

- Considerar la metodología propuesta como una referencia evaluar el efecto de la adición de atributos sintéticos a los datos sobre el desempeño de métodos de clasificación.
- Proponer otros tipos de atributos sintéticos, por ejemplo, con funciones exponenciales, trigonométricas o combinación de varias funciones no lineales.

Trabajo a futuro

Gracias a los resultados obtenidos en este artículo, se plantea la necesidad de continuar realizando pruebas con otros tipos de atributos sintéticos, para enfrentar de mejor manera las complejidades que se pueden encontrar en datos etiquetados. Se tiene planeado el uso de un algoritmo evolutivo para la generación de atributos sintéticos.

Conclusiones

Las complejidades de los datos degradan el desempeño de los métodos de aprendizaje supervisado. Una forma de enfrentar este problema es la generación de atributos sintéticos a los datos. En este artículo se analizó el desempeño de seis métodos de clasificación cuando se agregan atributos sintéticos.

Se realizaron experimentos para evaluar mediante la métrica F1-score, si la adición de atributos sintéticos de tipo polinomial puede ayudar a lograr un mejor desempeño en los clasificadores: SVM, Árbol de decisión, Regresión logística, Naive Bayes, clasificador K-NN y red neuronal multicapa. Además, se combinó este enfoque con la técnica de selección de atributos MrMR, para reducir la dimensionalidad mediante la eliminación de los atributos menos relevantes. Analizando los resultados se encontró que, en algunos conjuntos de datos, los clasificadores tienen un desempeño alto; con la selección y la adición de atributos sintéticos polinomiales se incrementa un poco el valor de F1-score. Así mismo, en los conjuntos de datos donde los clasificadores alcanzan valores bajos, la mejora en el desempeño de los clasificadores no es alta. Se requieren más estudios para determinar si es posible generar atributos de otro tipo que ayuden a lograr un mejor tratamiento a las complejidades presentes en los datos.

El código en Python desarrollando a lo largo de esta investigación se encuentra públicamente disponible en la siguiente liga: https://github.com/arthurp215/Clasificador-Polinomial-MrMR.git.

Referencias

[1] A. C. Lorena, A. I. Maciel, P. B. C. de Miranda, I. G. Costa, and R. B. C. Prudêncio, “Data complexity meta-features for regression problems,” *Mach Learn*, vol. 107, no. 1, pp. 209–246, Jan. 2018, doi: 10.1007/s10994-017-5681-1.

[2] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, 2020, doi: 10.1016/j.neucom.2019.10.118.

[3] S. Xiang, Y. Fu, G. You, and T. Liu, “Attribute analysis with synthetic dataset for person re-identification,” Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.07139

[4] J. Brownlee, “Imbalanced Classification with Python Choose Better Metrics, Balance Skewed Classes, and Apply Cost-Sensitive Learning,” 2020.

[5] A. J. Larner, *The 2x2 Matrix*. Springer International Publishing, 2021. doi: 10.1007/978-3-030-74920-0.
[6] H. Dalianis, *Clinical text mining: Secondary use of electronic patient records*. Springer International Publishing, 2018. doi: 10.1007/978-3-319-78503-5.
[7] A. H. Alsaffar, "Empirical study on the effect of using synthetic attributes on classification algorithms," *International Journal of Intelligent Computing and Cybernetics*, vol. 10, no. 2, pp. 111–129, 2017, doi: 10.1108/IJICC-08-2016-0029.
[8] Mazzanti Samuele, "'MRMR' Explained Exactly How You Wished Someone Explained to You | by Samuele Mazzanti | Towards Data Science," Feb. 12, 2021. https://towardsdatascience.com/mrmr-explained-exactly-how-you-wished-someone-explained-to-you-9cf4ed27458b (accessed Jul. 15, 2022).
[9] M. Billah and S. Waheed, "Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection," *Multimed Tools Appl*, vol. 79, no. 33–34, pp. 23633–23643, Sep. 2020, doi: 10.1007/s11042-020-09151-7.
[10] D. Dua and C. Graff, "UCI Machine Learning Repository." 2017. [Online]. Available: http://archive.ics.uci.edu/ml