

Recocido simulado para la optimización de parámetros SVM en la predicción del rendimiento de los estudiantes

Esraa A. Mahareek, Abeer S. Desuky, Habiba Abdullah El-Zhni

Facultad de Ciencias Departamento de Matemáticas, Universidad Al-Azhar, Egipto

Información del artículo

Historial del artículo:

Recibido el 29 de septiembre de 2020

Revisado el 17 de febrero de 2021

Aceptado el 8 de abril de 2021

Palabras clave:

Optimización bioinspirada

Datos educativos

Toma de decisiones educativas

Predicción

recocido simulado

rendimiento de los estudiantes

Optimización de parámetros SVM

educación universitaria

ABSTRACTO

La educación superior es una parte importante y crítica de la educación en todo el mundo. En el último año, el mundo se ha volcado cada vez más hacia la educación en línea debido al estallido de la pandemia de Covid-19; por lo tanto, mejorar este sistema educativo se convirtió en un asunto urgente. Los sistemas de aprendizaje en línea son un entorno primordial para la adquisición de datos educativos que pueden provenir de diferentes fuentes, especialmente instituciones académicas. Estos datos se pueden utilizar principalmente para analizar y extraer información útil para ayudar a comprender el desempeño de los estudiantes universitarios e identificar los factores que lo afectan. Para extraer información significativa de estos grandes volúmenes de datos, las organizaciones académicas deben extraer los datos con gran precisión. En este trabajo, se seleccionaron, preprocesaron, limpiaron y filtraron tres conjuntos de datos reales diferentes para aplicar la máquina de vectores de soporte (SVM) con kernel de perceptrón multicapa (kernel MLP) y optimizar sus parámetros utilizando el algoritmo de recocido simulado (SA) para mejorar el objetivo. valor de la función. Al examinar el espacio de búsqueda, SA tiene la ventaja de escapar de los mínimos locales ya que ofrece la posibilidad de aceptar al peor vecino como solución de manera controlada. Los resultados muestran que el sistema diseñado puede determinar los mejores parámetros de SVM utilizando SA y, por lo tanto, presenta una mejor evaluación del modelo.

Este es un artículo de acceso abierto bajo la [licencia CC BY-SA](#).



Autor correspondiente:

Esraa Alaa Mahareek

Facultad de Ciencias Departamento de Matemáticas

Universidad Al-Azhar, Egipto

Correo electrónico: esraa.mahareek@azhar.edu.eg

1. INTRODUCCIÓN

COVID-19 conocido como pandemia según fue declarado por la Organización Mundial de la Salud (OMS) en el último año. Esta pandemia interrumpió la educación en todo el mundo, ya que los cierres a nivel nacional obligaron a las instituciones a cerrar temporalmente. Se estima que los cierres afectaron a alrededor del 70% de la población estudiantil total en todo el mundo. Los algoritmos de minería de datos se usan ampliamente para descubrir patrones ocultos de datos para ayudar a los tomadores de decisiones, se convirtió en una herramienta eficiente para encontrar la información descubierta de los grandes datos. Al igual que las organizaciones comerciales [1], las universidades están operando hoy en entorno dinámico y fuertemente competitivo [2] y la educación hoy en día no se limita a la enseñanza en el aula, sino que se va a otras formas tales como sistemas de educación en línea, educación basada en la web, seminarios, aprendizaje basado en proyectos, talleres, etc. Minería de datos es muy importante en los sistemas educativos como se muestra en la Figura 1, pero todos estos sistemas no pueden tener éxito sin una evaluación precisa, por lo que, para tener un sistema educativo exitoso, se debe mantener un sistema de evaluación bien definido y preciso, la predicción del rendimiento de los estudiantes con alto la precisión es demasiado útil para seleccionar a los estudiantes con bajos niveles de rendimiento desde el principio de

proceso de aprendizaje. Las universidades modernas recopilan grandes volúmenes de datos de los estudiantes, que se utilizan para mejorar el proceso educativo.

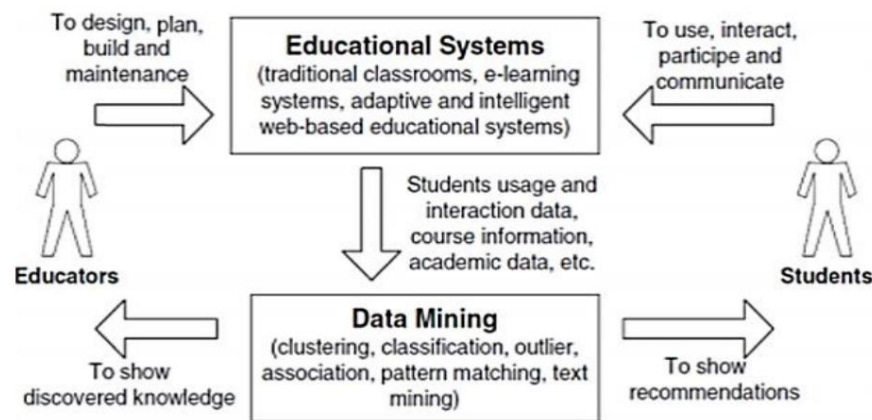


Figura 1. Ciclo de minería de datos para sistemas educativos [1]

Tanto la minería de datos como el aprendizaje automático utilizaron los mismos métodos. Pero hay una diferencia, el aprendizaje automático se centra en la predicción, en función de propiedades conocidas, mientras que la minería de datos se centra en la identificación de propiedades desconocidas. La máquina de vectores de soporte (SVM) es una técnica de aprendizaje automático que construye un clasificador binario lineal. Define el límite de decisión entre dos clases [2]. La optimización es el proceso de lograr la mejor solución para un problema, hay muchos algoritmos de optimización como el algoritmo SPSS estándar [3], [4] que se usa para optimizar sistemas con múltiples parámetros desconocidos, Descenso de gradiente que se usa para encontrar un mínimo local de una función diferenciable [5], y recocido simulado (SA) que se utiliza para aproximar el óptimo global de una función dada en un gran espacio de búsqueda [6]. Por lo tanto, fue elegido para este trabajo.

SA es un popular algoritmo de optimización inspirado en la base del recocido de metales fundidos (enfriamiento lento después del calentamiento) para cristalizar sus estructuras [7], fue inventado en 1983 por Kirkpatrick *et al.* y ellos y también algunos otros investigadores demostraron analíticamente que SA puede escapar de los óptimos locales y converger al óptimo global. Un grupo de investigadores estudió datos de estudiantes durante la última década para predecir el rendimiento de los estudiantes. Los enfoques de minería de datos y análisis de correlación cada uno de estos enfoques generan diferentes niveles de éxito. VK Pal y Vimal Kamlesh Kumar Bhatt [8] propusieron investigar el primer conjunto de datos aplicando la red neuronal artificial (aprendizaje profundo) después de dividir los datos en dos subconjuntos, el conjunto de entrenamiento que contiene el 70 % de los datos originales y el conjunto de prueba que contiene el 30 % restante. La precisión resultante para el conjunto de prueba es del 97,749 % y la tasa de error correspondiente es del 2,251 %. YK Salal, SM Abdullaev y Mukesh Kumar [9] también propusieron investigaciones para construir modelos de clasificación para el mismo conjunto de datos e implementar algoritmos como NaiveBayes con una precisión del 73,1895 %, árbol de decisión (J48) con una precisión del 76,57 %, Randomtree con una precisión del 67,95 %, REPTree con precisión 76,73 %, JRp con precisión 74,11 %, OneR con precisión 76,73 %, simplelogistic con precisión 73,65 % y ZeroR con precisión 30,97 %. Después de implementar estos algoritmos en el conjunto de datos de rendimiento de los estudiantes, compara el resultado de la implementación con el mejor modelo en el proceso de predicción.

Otra investigación propuso un método de clasificación basado en un algoritmo PSO meta-heurístico para predecir el resultado final de los estudiantes según sus actividades y los resultados mejoraron en un 89% [10]. D. Kabakchieva [11] también propuso un algoritmo para la clasificación mediante la aplicación de cuatro clasificadores diferentes: OneR Rule Learner, Neural Network, Decision Tree y K-Nearest Neighbour, y la red neuronal logró la mayor precisión de clasificación con un 73,9 %, seguida del 72,74 % para Decision Tree y 70,49% para el modelo k-NN. S. Hussain, Neama Abdulaziz Dahan, Fadl Mutaheer Ba-Alwi y Najoua Ribata [12] utilizaron algoritmos de clasificación en WEKA y aplicaron la selección de características para seleccionar 12 de 33 atributos para predecir el rendimiento de los estudiantes.

La optimización es el proceso de lograr la mejor solución para un problema (parámetros SVM) en este artículo utilizando la técnica de optimización SA que ayuda a mejorar el valor de la función objetivo (precisión de clasificación) al evitar los mínimos locales y presenta un estudio comparativo para el rendimiento académico de los estudiantes, cada de estos algoritmos se compara en función de su precisión para identificar el modelo más apropiado para este trabajo.

La comparación de nuestros resultados en la sección 7 con trabajos publicados anteriormente muestra claramente que nuestra propuesta SVM-SA

da mejores resultados de exactitud, precisión, sensibilidad y medidas f que mejoran la predicción del rendimiento académico de los estudiantes para los tomadores de decisiones. Este documento está organizado de la siguiente manera: en la sección 2 y 3 se explica tanto el clasificador SVM como el algoritmo SA, en la sección 4 se describe el modelo SVM-SA propuesto, la sección 5 describe los datos utilizados, la sección 6 mide las medidas de evaluación, la sección 7 y 8 el resultados y conclusiones.

2. MÁQUINA DE VECTORES DE SOPORTE

En 1995, SVM se desarrolló originalmente utilizando el principio de minimización de riesgos estructurales y la teoría de Vapnik-Chervonenkis, es una técnica de aprendizaje automático supervisado que se utiliza tanto para problemas de clasificación como de regresión. Las SVM se usan más comúnmente en problemas de clasificación porque tienen un alto rendimiento y capacidad de generalización.

Las SVM se basan principalmente en la idea de encontrar el mejor hiperplano que maximice el margen (distancia a los puntos más cercanos) entre los puntos de datos +ve y -ve más cercanos [13], el límite de clase para datos linealmente separables, lo que brinda una mayor posibilidad de nuevos los datos se clasifican correctamente [14], suponga que los datos de entrenamiento x_i pertenecen a la clase +ve o -ve, y_i es el valor de la etiqueta de salida correspondiente a la clase de elemento x_i , la formulación original del algoritmo SVM busca una decisión lineal superficie utilizando la fórmula $(x, w) + b = 0$ donde w es un vector de coeficiente dimensional y b es el desplazamiento [15]. La SVM lineal logra un hiperplano óptimo al resolver el siguiente problema de optimización:

$$\min_{w, b} \frac{1}{2} \|w\|^2 + \sum_{i=1}^n y_i (w \cdot x_i + b) \quad (1)$$

Este problema de optimización cuadrática se puede resolver encontrando el punto de silla de la función Lagrangiana:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^n y_i (w \cdot x_i + b) - \sum_{i=1}^n \alpha_i [y_i (w \cdot x_i + b) - 1] \quad (2)$$

Donde son variables de Lagrange, después de aplicar las condiciones KKT para un máximo de (2) se obtienen poniendo a cero el gradiente de Lagrange con respecto a las variables primarias w y b y escribiendo las condiciones complementarias [16]:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3)$$

$$b = \sum_{i=1}^n \alpha_i y_i \quad (4)$$

$$\alpha_i \geq 0, \quad [y_i (w \cdot x_i + b) - 1] = 0 \quad (5)$$

Por (3), la solución del vector de peso w del problema SVM es una combinación lineal del conjunto de vectores x_1, \dots, x_n entrenamiento. De acuerdo a condiciones complementarias (6) w depende del vector al que corresponde el $\alpha_i = 0$. Los cuales se denominan vectores soporte que definen completamente el hiperplano de margen máximo, luego de sustituir (3) y (4) en (2) la forma dual Lagrangiana $L(\alpha)$ de (2) se obtiene de la siguiente manera:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (6)$$

En (7), (8) y (9) se presentan las funciones de núcleo polinomial, núcleo sigmoide y base radial, respectivamente. Estas funciones se usan para encontrar el hiperplano óptimo, en esta propuesta usamos el núcleo sigmoideo (MLP) que también se llama ANN de alimentación hacia adelante con tres capas de neuronas, cada neurona usa una función no lineal para la activación excepto la de entrada y también aplica el concepto de retropropagación durante el entrenamiento en red [17]. El peso, el sesgo y son los parámetros de configuración del perceptrón multicapa (MLP) [18].

$$\text{Núcleo polinomial: } (x, x') = (1 + x \cdot x')^k \quad (7)$$

$$\text{Núcleo sigmoideo: } (x, x') = \tanh(x \cdot x') \quad (8)$$

donde la constante de intersección

$$\text{Núcleo de función de base radial (RBF): } (x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \quad (9)$$

donde esta el parametro del kernel

Hay dos problemas en el procedimiento de optimización del clasificador SVM [13]: 1) Cómo seleccionar características relevantes y filtrar características irrelevantes para construir el clasificador SVM; 2) Cómo ajustar correctamente el parámetro de penalización C y los parámetros del hiperplano [19]. Los parámetros de SVM, como los parámetros del kernel y el parámetro de penalización, tienen una gran influencia en la precisión y complejidad de los modelos de clasificación. se propusieron numerosos algoritmos de optimización evolutiva para optimizar las SVM; En este documento, SA se propone como un algoritmo de optimización que sigue una estrategia de búsqueda que mejora el valor de la función objetivo para encontrar la mejor configuración de parámetros que pueda mejorar en gran medida el rendimiento del clasificador SVM.

3. ALGORITMO DE RECOCIDO SIMULADO

El algoritmo SA es un método de búsqueda local inventado para evitar los mínimos locales [7], [20]. La principal ventaja de SA en comparación con los métodos de optimización más antiguos es su capacidad para escapar de los mínimos locales. Este método se basa principalmente en elegir un movimiento al azar en cada etapa en lugar de seleccionar el mejor movimiento (mejor vecino) entre los movimientos disponibles, si el nuevo estado mejoró (redujo) el costo, se acepta como el siguiente estado mientras que si causó el incremento de costo, se acepta solo con una probabilidad P. P recibe el nombre de probabilidad de Metropolis y se define como:

$$(\ddot{y}) = \frac{\ddot{y}}{T} \quad (10)$$

Donde $\ddot{y}E$ representa el cambio de energía (valor de la función de costo) causado por el cambio de estado T es la temperatura o variable similar a la temperatura que controla esta probabilidad. Existe una "función generativa" que denota la forma de actualizar las variables en cada intento y de hecho es la función que especifica la velocidad de convergencia. En SA típico, la función generativa es una función gaussiana o de Boltzmann:

$$(\ddot{y}) = (2\pi)^{-\frac{D}{2}} \exp\left(-\frac{\ddot{y}^2}{2}\right) \quad (11)$$

donde D es una dimensión del espacio de búsqueda (número de variables en la función de costo). $\ddot{y}X$ muestra la tasa de cambio de X (vector de variables). Entonces, $X=X_0+\ddot{y}X$ donde X_0 es el estado actual y X el siguiente estado de las variables. La temperatura en la k-ésima etapa del algoritmo se puede encontrar usando (12). Los pasos del algoritmo SA [21] mostrado en la figura 2.

$$= \frac{0}{\text{en}} \quad (12)$$

1. Choosing an initial solution i from the group of feasible solution S.
2. Choosing the initial temperature $T_0 > 0$
3. Selecting the number of iterations $N(t)$ at each temperature
4. Selecting the final temperature T_F
5. Determining the process of the temperature reduction until it reaches T_F .
6. Setting the temperature exchange counter n to zero for each temperature
7. Creating the j solution at the neighbourhood of the i solution
8. Evaluation of the objective function at any temperature and calculate $\Delta = z(j) - z(i)$
9. Accepting the solution j , if $\Delta < 0$. Else, generation a random number then select solution j
10. Setting $n=n+1$. If n is equal to $N(t)$ then go to 12. Otherwise, go to 7.
11. Reducing the temperature. If it reaches T_F then stop. Otherwise, go to 6.

Figura 2. Algoritmo SA

4. MÉTODO DE INVESTIGACIÓN

La penalización y el kernel son parámetros SVM con un gran impacto en la precisión y complejidad del modelo de clasificación. Este artículo propone una nueva evolución para el modelo SVM utilizando SVM con kernel MLP y emplea SA para optimizar sus parámetros que se expresan como $P1$ para la pendiente o peso y $P2$ para la constante de intersección o sesgo donde $P1 > 0$ y $P2 < 0$. Por morosidad, el valor se establece en 1 y -1; por lo tanto, se puede disminuir el error de clasificación. En esta sección, describimos el modelo SA-SVM propuesto como se muestra en la Figura 3 para encontrar los valores óptimos de los parámetros de SVM.

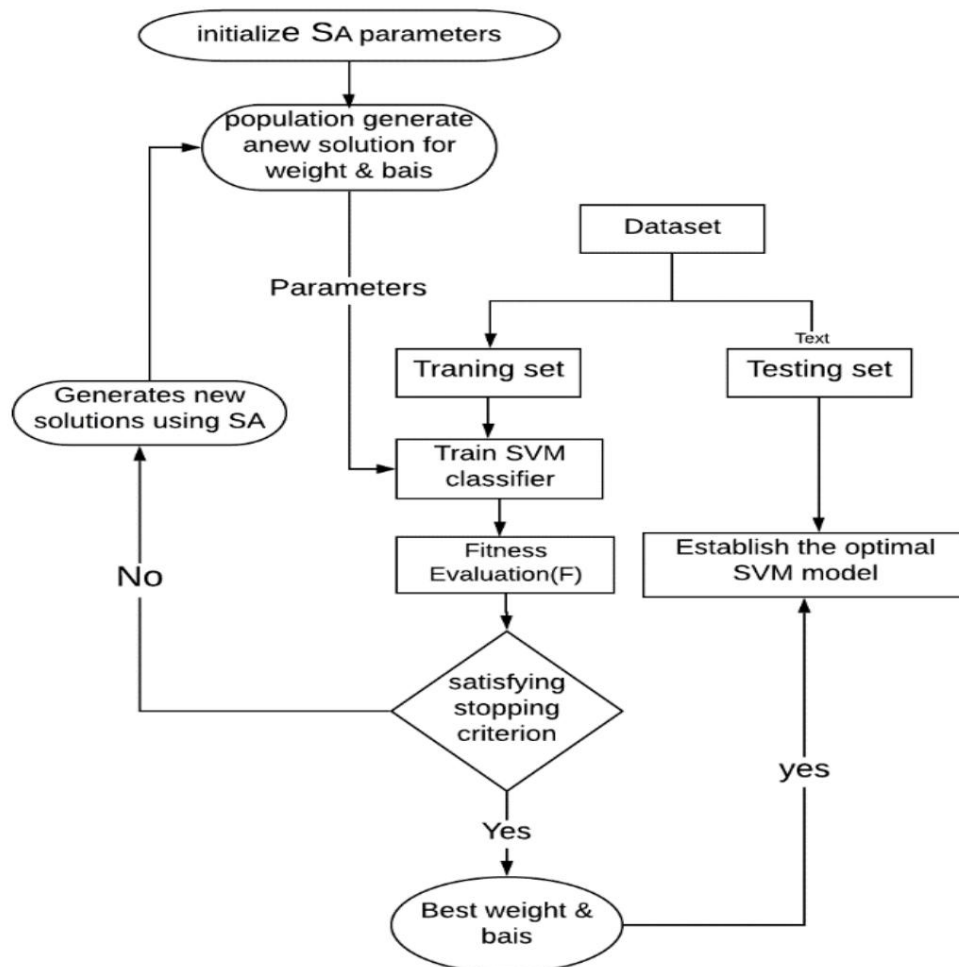


Figura 3. Diagrama de flujo del modelo propuesto (SA-SVM)

Los pasos principales en el algoritmo SA son: 1) generar vecino; 2) evaluar la función objetivo (precisión de clasificación); 3) asignar una temperatura inicial; 4) cambiar la temperatura; 5) esquemas de enfriamiento, y 6) parada [21]. La solución inicial es uno de los componentes importantes de SA que se generó aleatoriamente seleccionado entre un espacio de solución factible en este documento. La solución inicial en nuestro algoritmo está representada por un vector P de dos elementos como (13). Mientras que P_1 se asigna al peso, P_2 se asigna al sesgo.

$$P = [P_1, P_2] = [w, b] \quad (13)$$

$[P_1, P_2]$ es un vector que especifica los parámetros del kernel MLP. El núcleo MLP toma la forma:

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad \text{donde } \gamma > 0 \text{ y } P_2 < 0. \text{ El valor predeterminado es } [1, -1] \quad (14)$$

Se selecciona aleatoriamente una solución factible para que sea una solución inicial. La función objetivo es un factor importante del que depende SA durante su desempeño para evaluar las soluciones individuales. Formulamos la función objetivo para que dependa principalmente de la precisión de clasificación de SVM representada por la solución dada. La precisión con la que se clasifican los datos de entrenamiento cuando la clasificación se realiza utilizando los parámetros presentados por una solución sirve como costo para una solución determinada. El costo $Z(P)$ para una solución P se calcula sobre el conjunto de datos de entrenamiento (con un tamaño de N) usando (15).

$$Z(P) = \frac{\text{\#Instancias verdaderamente clasificadas}}{N} \quad (15)$$

La temperatura inicial también tiene una gran importancia ya que un parámetro tiene un gran efecto sobre la posibilidad de seleccionar una mala solución. Entonces, si la temperatura inicial tiene un valor alto, una solución con un mal valor de la función objetivo puede tener una alta probabilidad de ser aceptada. Si bien se considera un valor bajo para el valor inicial

temperatura aumenta la probabilidad de que la solución sea un óptimo local, en este trabajo se elige que las temperaturas iniciales estén en un rango de 0:500.

5. PREPROCESAMIENTO DE DATOS

Los datos de entrada y salida se preprocesaron limpiando los valores faltantes, convirtiendo los datos nominales en datos numéricos, convirtiendo los datos en clase binaria (0 significa falla, 1 significa éxito) y dividiendo el conjunto de datos en dos partes: conjuntos de datos de entrenamiento y prueba (proporción de 70 %: 30 %) sin ninguna selección de características para ningún conjunto de datos, no utilizamos la validación cruzada para que la comparación fuera justa porque los artículos utilizados en la comparación utilizaron la misma proporción de datos de entrenamiento y prueba.

6. DESCRIPCIÓN DE DATOS

El primer conjunto de datos consta de 649 instancias diferentes con 33 atributos diferentes, el conjunto de datos de rendimiento de este estudiante se recopila de dos escuelas secundarias portuguesas (Gabriel Pereira (GP) y Mousinho da Silveira (MS)). El conjunto de datos contiene atributos para estudiantes como calificaciones académicas, atributos sociales, atributos demográficos y atributos relacionados con la escuela. Los datos se recopilaron de los estudiantes utilizando los informes y cuestionarios escolares. Los detalles del conjunto de datos se muestran en la Tabla 1 [9] [22]. El segundo conjunto de datos es de tres universidades diferentes, Duliajan, Doomdooa y Digboi College of Assam, India. Inicialmente, se recopilaron datos de veintidós atributos [12]. El tercer conjunto de datos proviene de la entrada común realizada por la Universidad de Dibrugarh. Los datos recopilados con 12 atributos fueron de estudiantes que acudieron para recibir asesoramiento y admisión en las facultades de medicina de Assam en el año 2013 [23]. Los tres conjuntos de datos están desequilibrados, esto se debe a la baja tasa de repetición entre los estudiantes en los lugares de compilación de la base de datos según la Tabla 2.

Tabla 1. Información de conjuntos de datos

conjunto de datos	Instancias	Nº de atributos
curso de portugués	649 300	33
archivo de savia	666	22
datos CEE		12

Tabla 2. Información sobre el número de clases

conjunto de datos	pase	No pasar
curso de portugués	452	197
archivo de savia	224	76
datos CEE	509	157

7. EVALUACIÓN DEL MODELO

La precisión de la clasificación siempre aprovecha el primer vistazo cuando se construye un modelo para un problema de clasificación como el número de instancias predichas correctamente de todas las predicciones realizadas, pero la precisión de la clasificación no es suficiente por sí sola para evaluar un modelo, especialmente en el caso de una clasificación de datos desequilibrada. Por lo tanto, consideramos algunas otras medidas, como la sensibilidad, la precisión y las medidas F [24]. Las ecuaciones de medición utilizadas en la evaluación del modelo enumeradas en la Tabla 3 fueron: TP para verdaderos positivos, TN para verdaderos negativos, FP para falsos positivos y FN para falsos negativos [25].

Tabla 3. Ecuaciones de medida

La medida	Fórmula
Precisión, tasa de reconocimiento	$\frac{TP}{TP + FP}$
sensibilidad, tasa de verdaderos positivos, recuerdo	$\frac{TP}{TP + FN}$
Precisión	$\frac{TP}{TP + FP}$
medidas F	$\frac{2 \times \text{Precisión} \times \text{sensibilidad}}{\text{Precisión} + \text{sensibilidad}}$

8. RESULTADOS

La plataforma adoptada para desarrollar el algoritmo SA-SVM es una laptop con las siguientes características: Intel(R) Core(TM) i7-4600 CPU@2.10GHz, 8G RAM, un Windows 10 pro como sistema operativo usando

Versión de software MATLAB R2015a. Para evaluar el algoritmo propuesto, se utilizaron tres conjuntos de datos de clasificación estándar. Los conjuntos de datos se obtuvieron del Repositorio de aprendizaje automático de la Universidad de California en Irvin (UCI), las propiedades de todos los conjuntos de datos se enumeran en la Tabla 1, cada conjunto de datos se divide en dos partes en las que los conjuntos de datos de entrenamiento y prueba tienen una proporción de 70% y 30% respectivamente y los resultados se muestran en la Tabla 4, Tabla 5 y Tabla 6.

Tabla 4. Resultados del conjunto de datos de CEE

Conjunto de datos de la CEE	Precisión	Sensibilidad	61,8% 52,8%	Precisión	medidas F
SVM (núcleo MLP)				58%	55,29%
SVM (núcleo MLP) y SA	69,34%		75,28%	63,2%	68,71%

Tabla 5. Resultados del conjunto de datos del curso de portugués

Conjunto de datos del curso de portugués	Exactitud	Sensibilidad	Precisión SVM (núcleo MLP)	78,35 %	medidas F
76,21 % 97,65 % SVM (núcleo MLP) &SA	90,72 %	97,56 %	91,95 %		85,6%
					94,67%

Tabla 6. Resultados del conjunto de datos Sapfile

Conjunto de datos Sapfile	Precisión	Sensibilidad	61,1% 55,22%	Precisión	medidas F
SVM (núcleo MLP)	67,77%	88%		88%	67,88%
SVM (núcleo MLP) y SA				73,75%	80,27%

De las tablas 4 a 8, está claro que nuestro método propuesto muestra un rendimiento comparativo sin selección de características con todos los demás algoritmos de clasificación en términos de precisión de predicción. El conjunto de datos del curso portugués mostró una precisión del 78,35 % para SVM con kernel MLP, esta precisión aumentó al 90,72 % después de aplicar el método propuesto, así como la sensibilidad y las medidas f, como se muestra en la Tabla 5, y es mejor que la precisión de otros clasificadores presentados, como se muestra en Tabla 7. El conjunto de datos de CEE mostró una precisión del 61,8 % para SVM con kernel MLP, esta precisión aumentó al 69,34 % después de aplicar el método propuesto, así como la sensibilidad y las medidas f, como se muestra en la Tabla 4, y es mejor que NaiveBayes y el árbol de decisiones (J48), ZeroR, REPTree, OneR, RandomTree, JRip y SimpleLogistic, como se muestra en la Tabla 8.

El conjunto de datos Sapfile mostró una precisión del 61,1 % para SVM con kernel MLP, esta precisión aumentó al 67,77 % después de aplicar el método propuesto, así como la sensibilidad y las medidas f, como se muestra en la Tabla 6, y es mejor que la precisión de BayesNet, como se muestra en la Tabla 9. También , la gran mejora en las medidas f para los tres conjuntos de datos utilizados que alcanzaron el 13 % demuestra contundentemente la eficiencia de nuestro modelo SVM-SA propuesto para abordar el problema del desequilibrio en los datos. Todo lo anterior confirma la efectividad del método propuesto.

Tabla 7. Comparación de conjuntos de datos de cursos de portugués

Método	Precisión Ref.	rango de Wilcoxon
SVM (núcleo MLP) y SA	90,72 % propuesto 1	
NaiveBayes	68,25 % [9] 6	
Árbol de decisión	67,79 % [9] 7	
Árbol aleatorio	53,46 % [9] 8	
Árbol REP	75,19 % [9] 3	
JRip	70,72% [9] 5	
UnoR	76,73 % [9] 2	
SimpleLogistic	71,34 % [9] 4	
ZeroR	30,97 % [9] 9	

Tabla 8. Comparación de conjuntos de datos de CEE

Método	Precisión	Árbol.	rango de Wilcoxon
SVM (núcleo MLP) y SA	69,34 %	Propuesto	1
Árbol de decisión (J48)	64,71 %	[23] [23]	2
bayesiana ingenua	57,81 %		4
CeroR	31,53 %	Poner	9
REPÁrbol	55,55 %	Poner	5
SimpleLogística	60,36 %	Poner	3
Raridad	51,05 %	Poner	7
JRip	54,80 %	Poner	6
Árbol aleatorio	33,5 %	Poner	8

Tabla 9. Comparación de conjuntos de datos Sapfile

Método	Precisión	Árbitro.	rango de Wilcoxon
SVM (núcleo MLP) y SA	67,77%	propuesto	1
BayesNet	65,33%	[12]	2
bayesiana ingenua	51,11%	Poner	5
CeroR	41,66%	Poner	8
REPÁrbol	55%	Poner	4
SimpleLogística	63,33%	Poner	3
Raridad	50,66%	Poner	6
JRip	47,77%	Poner	7
Árbol aleatorio	37,61%	Poner	9

9. CONCLUSIÓN

Las técnicas de aprendizaje automático con datos educativos se pueden utilizar para mejorar el proceso de aprendizaje de los estudiantes en los institutos de educación superior. Los investigadores desarrollaron diferentes métodos para predecir el rendimiento de los estudiantes en los cursos matriculados, para proporcionar información valiosa que ayude a facilitar la retención de los estudiantes en esos cursos. Los instructores pueden utilizar esta información para identificar de manera temprana a los estudiantes que podrían necesitar ayuda en su estudio. En nuestro trabajo, SVM se aplicó en tres conjuntos de datos reales diferentes, luego, se utilizó una hibernación entre SVM con kernel MLP y SA para mejorar los resultados y, finalmente, compararlos con los resultados de otros algoritmos. Los resultados mostraron que el método propuesto se volvió mejor después de aplicar la técnica de optimización SA y presenta mayor rendimiento que otros métodos.

REFERENCIAS

- [1] A. Soni, Vivek Kuma, RK y D. Hemavathi, "Predicción del desempeño de los estudiantes mediante técnicas de minería de datos", *Revista Internacional de Matemáticas Puras y Aplicadas*, vol. 119, núm. 12, págs. 221-227, 2018.
- [2] A. Kaur, Nitin Umesh y Barjinder Singh, "Enfoque de aprendizaje automático para predecir el rendimiento académico de los estudiantes", *Revista internacional de investigación en ciencia aplicada y tecnología de ingeniería (IJRASET)*, vol. 6, núm. 4, págs. 734-742, abril de 2018.
- [3] MA Ahmad, Shun-ichi Azuma y Toshiharu Sugie, "Identificación de sistemas Hammerstein de tiempo continuo mediante aproximación estocástica de perturbaciones simultáneas", *Sistemas expertos con aplicaciones*, vol. 43, págs. 51-58, enero de 2016, doi: 10.1016/j.eswa.2015.08.041.
- [4] NSA Shukor, MA Ahmad y MZM Tumari, "Ajuste de PID basado en datos basado en dinámicas de experimentación seguras para el control de salpicaduras de líquidos", *2017 IEEE 8th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malasia, 2017, págs. 62-66, doi: 10.1109/ICSGRC.2017.8070569.
- [5] SS Du, Jason D. Lee, Haochuan L, Liwei Wang y Xiyu Zhai, "Gradient Descent Finds Global Minima of Deep Neural Networks", en *Conferencia internacional sobre aprendizaje automático*, Long Beach, California, vol. 97, págs. 1675-1685, 2019.
- [6] AS Ramli, Muhammad Ikram Mohd Rashid y Mohd Ashraf Ahmad, "Estrategia de gestión de energía de HEV basada en recocido simulado", *Revista internacional de ingeniería integrada*, vol. 12, núm. 2, págs. 30-37, febrero de 2020.
- [7] M. Vakil-Baghmisheh y A. Navarbarf, "Un algoritmo de recocido simulado muy rápido modificado", *Simposio internacional sobre telecomunicaciones de 2008*, Teherán, Irán, 2008, págs. 61-66, doi: 10.1109/ISTEL.2008.4651272.
- [8] VK Pal y Vimal Kamlesh Kumar Bhatt, "Predicción de rendimiento para estudiantes de posgrado que utilizan redes neuronales artificiales", *Revista internacional de tecnología innovadora y exploración de ingeniería (IJITEE)*, vol. 8, núm. 7S2, mayo de 2019.
- [9] YK Salal, SM Abdullaev y Mukesh Kumar, "Minería de datos educativos: predicción del rendimiento de los estudiantes en el ámbito académico", *Revista internacional de ingeniería y tecnología avanzada (IJEA)*, vol. 8, núm. 4C, págs. 54-59, abril de 2019.
- [10] SM-. H. Hasheminejad y M. Sarvmili, "S3PSO: Predicción del rendimiento de los estudiantes basada en la optimización del enjambre de partículas", *Journal of AI and Data Mining*, vol. 7, núm. 1, págs. 77-96, 2019, doi: 10.22044/JADM.2018.5506.1662.
- [11] D. Kabakchieva, "Predicción del rendimiento de los estudiantes mediante el uso de algoritmos de clasificación de minería de datos", *International Revista de Ciencias de la Computación e Investigación de Gestión*, vol. 1, no. 4, 2012.
- [12] S. Hussain, Neama Abdulaziz Dahan, Fadl Mutaheer Ba-Alwi y Najoua Ribata, "Extracción de datos educativos y análisis del rendimiento académico de los estudiantes mediante WEKA", *Revista de ingeniería eléctrica e informática de Indonesia*, vol. 9, núm. 2, págs. 447-459, febrero de 2018.
- [13] F. Ye, "Evolución del modelo SVM basado en un método híbrido que usa técnicas de optimización de enjambre en combinación con un algoritmo genético para el diagnóstico médico", *Springer Science+Business Media New York*, vol. 77, núm. 3, págs. 3889-3928, 2016.
- [14] F. Albu, Adelaida MATEESCU y Neculai DUMITRIU, "Selección de arquitectura para una red de alimentación directa multicapa", en *Conferencia internacional sobre microelectrónica e informática*, págs. 131-134, octubre de 1997.
- [15] F. Albu y D. Martinez, "La aplicación de máquinas de vectores de soporte con núcleos gaussianos para superar co-

- interferencia de canal", *Neural Networks for Signal Processing IX: Proceedings of the IEEE Signal Processing Society Workshop de 1999 (Cat. No.98TH8468)*, Madison, WI, EE. UU., 1999, págs. 49-57, doi: 10.1109/NNSP.1999.788122.
- [16] F. Albu, A. Mateescu, JCM Mota y B. Dorizzi, "Ecuación de canales adaptativos mediante redes neuronales", *Actas de ITS'98. Simposio Internacional de Telecomunicaciones SBT/IEEE (Cat. No.98EX202)*, Sao Paulo, Brasil, 1998, pp. 438-441 vol.2, doi: 10.1109/ITS.1998.718433.
- [17] KAA Kamarulzaini, N. Ismail, MHF Rahiman, MN Taib, NAM Ali y SN Tajuddin, "Evaluación de RBF y MLP en los parámetros sintonizados del kernel SVM para la clasificación de la calidad del aceite de madera de agar", *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malasia, 2018, págs. 250-254, doi: 10.1109/CSPA.2018.8368721.
- [18] D. Kaya, "Optimización de los parámetros de SVM con algoritmos híbridos CS-PSO para la enfermedad de Parkinson en el entorno de LabVIEW", *Hindawi*, 2019, doi: <https://doi.org/10.1155/2019/2513053>
- [19] Fuerza multinacional b. Ariffin, Nurlaila Ismail, Nor Azah Mohd Ali, Mohd Hezri Fazalul Rahiman, Saiful Nizam Tajuddin y Mohd Nasir Taib, "Perceptrón multicapa (MLP) tuned Kernel Parameter in Support Vector Machine (SVM) for Agarwood Oil Compounds Quality Classification", *Journal of Tomography Aplicación de sistemas y sensores*, vol. 1, no. 1 de junio de 2018.
- [20] . Kirkpatrick, CD Gelatt y MP Vecchi, "Optimización mediante recocido simulado", *ciencia*, vol. 220, núm. 4598, págs. 671-680, 1983.
- [21] AA Taleizadeh, Seyed Taghi Akhavan Niaki y Gholamreza Jalali Naini, "Optimización de sistemas de control de inventario de múltiples restricciones y productos múltiples con duración de período estocástico y orden de emergencia", *Journal of Uncertain Systems*, vol. 7, núm. 1, págs. 58-71, 2013.
- [22] Mahsa Afsharizadeh, Hossein Ebrahimpour-Komleh, "Uso de técnicas de minería de datos para predecir el consumo de alcohol en escuelas secundarias portuguesas", *1.ª Conferencia internacional sobre nuevos logros de investigación en ingeniería eléctrica e informática*, 2016.
- [23] S. Hussain, Rasha Atallah, Amirrudin Kamsin y Jiten Hazarika, "Clasificación, agrupamiento y minería de reglas de asociación en conjuntos de datos educativos utilizando herramientas de minería de datos: un estudio de caso", *Conferencia en línea de informática*. Springer, págs. 196-211, 2019.
- [24] . K. Karthikeyan y P. Kavipriya, "Sobre la mejora de la predicción del rendimiento de los estudiantes en los sistemas educativos mediante técnicas mejoradas de minería de datos", *Revista internacional de investigación avanzada en informática e ingeniería de software*, vol. 7, núm. 5, págs. 935-941, mayo de 2017, doi: 10.23956/ijarsse/SV7I5/0348.
- [25] B. Sekeroglu, Kamil Dimililer y Kubra Tuncal, "Predicción y clasificación del rendimiento de los estudiantes mediante algoritmos de aprendizaje automático", *8.ª Conferencia internacional sobre tecnología de la información y la educación*, págs. 7–11 de marzo de 2019, doi: <https://doi.org/10.1145/3318396.3318419>.