

Aprendizaje supervisado a través del alumno de clasificación

Técnicas para el Sistema Predictivo de Personal y

Actitudes sociales de los estudiantes de ingeniería

Omar Chamorro-Atalaya¹, Soledad Olivares-Zegarra²
Facultad de Ingeniería y Gestión
Universidad Nacional Tecnológica de Lima Sur (UNTELS)
Lima- Perú

Alejandro Paredes-Soria³, Óscar Samanamud-Loyola⁴,
Marco Anton-De los Santos⁵, Juan Anton-De los
santos⁶, Maritte Fierro-Bravo⁷
Facultad de Ciencias Económicas
Universidad Nacional Federico Villarreal
Lima-Perú

Victor Villanueva-Acosta⁸
Facultad de Ciencias Humanas
Universidad Autónoma del Perú
Lima-Perú

Resumen—En este escenario competitivo del sistema educativo, las instituciones de educación superior utilizan herramientas y técnicas de aprendizaje inteligente para predecir los factores del rendimiento académico de los estudiantes. Ante esto, el artículo tiene como objetivo determinar el modelo de aprendizaje supervisado para el sistema predictivo de actitudes personales y sociales de estudiantes universitarios de carreras profesionales de ingeniería. Para ello se utiliza la técnica Machine Learning Classification Learner por medio del software Matlab R2021a. Los resultados reflejan un sistema predictivo capaz de clasificar las cuatro clases de satisfacción (1: insatisfecho, 2: poco satisfecho, 3: satisfecho y 4: muy satisfecho) con una exactitud del 91,96%, una precisión del 79,09%, una Sensibilidad del 75,66 % y una Especificidad del 92,09%, en cuanto a la percepción de los estudiantes sobre sus actitudes personales y sociales. Como resultado, la institución superior podrá tomar medidas para monitorear y corregir las fortalezas y debilidades de cada variable relacionada con la satisfacción con la calidad del servicio educativo.

Palabras clave—Aprendizaje supervisado; estudiante de clasificación; sistema predictivo; actitudes personales y sociales; estudiantes de ingeniería

I. INTRODUCCIÓN

En la actualidad, la búsqueda permanente de la calidad educativa es uno de los principales objetivos de este sector [1], por lo que se han implementado estrategias y metodologías diseñadas para optimizar los factores de satisfacción de los estudiantes [2], [3]. Específicamente, en el ámbito de la educación superior universitaria, se evidencia la necesidad de cumplir con estándares de calidad en el servicio educativo ofrecido [4].

Por lo tanto, es necesaria la autoevaluación continua de las dimensiones relacionadas con la satisfacción de los estudiantes, para que esta mejore los procesos institucionales identificando sus fortalezas y debilidades [5]-[8]. Teniendo en cuenta los diversos factores que influyen, es complejo determinar las acciones y decisiones estratégicas que optimicen correctamente estos factores [9].

Uno de los factores que se relacionan con la satisfacción de los estudiantes es la autopercepción de las actitudes personales y sociales de los estudiantes universitarios. Como se indica en [10] existe la necesidad de evaluar la autopercepción de las actitudes personales y sociales en los estudiantes, ya que esto generará la obligación de reflexionar sobre cómo la comunidad universitaria está contribuyendo al desarrollo social. La importancia de identificar la autopercepción de la actitud personal y social del universitario radica en identificar qué aspectos necesitan ser mejorados, de tal forma que esto conduzca a un aporte significativo a la educación superior en el contexto social, permitiendo al estudiante cumplir con sus objetivos profesionales y personales [11], [12].

Como se indica en [13], el nivel educativo superior, tiene el deber de inculcar en los estudiantes universitarios el sentido de responsabilidad social propio del desempeño de la actividad profesional. En este sentido, la consecuencia de la adquisición de actitudes personales y sociales es un proceso que abarca una parte importante de la vida de los estudiantes en su paso por las diferentes organizaciones. Ahora bien, si desarrollamos este análisis en el contexto de la virtualización educativa, nos lleva a todos a preguntarnos en qué medida este escenario ha alterado aspectos relacionados con las actitudes personales y sociales de los estudiantes universitarios.

[14], [15] teniendo en cuenta. Como se indica en [16] que los estudiantes de ingeniería han percibido un mayor cambio en el proceso de enseñanza, debido al carácter teórico y práctico de las materias del plan de estudios.

Ante esto, a lo largo del tiempo, los sistemas de tecnologías de la información han sido diseñados para diferentes sectores organizacionales, desde el transaccional hasta el de toma de decisiones [17]. Dentro de estos sistemas de tecnologías de la información se encuentra la herramienta de minería de datos. La minería de datos es propicia para el tratamiento de grandes cantidades de información, y su propósito es generar conocimiento [18], [19]. La minería de datos utiliza bases de datos, de las cuales se extrae información de forma automatizada y mediante análisis matemáticos y estadísticos se deducen patrones y tendencias [20].

La minería facilita el análisis de datos, en comparación con la exploración tradicional que, debido a la gran cantidad de datos, hace que este proceso sea mucho más complejo [21].

Respecto al sector educativo, la Minería de Datos Educativos (EDM) es una disciplina emergente que busca desarrollar métodos para explorar los datos generados en el sector educativo, con el fin de lograr una mejor comprensión de las características de los estudiantes y la forma en que aprenden [22]. Su desarrollo utiliza técnicas estadísticas e inteligencia artificial para detectar patrones y anomalías en grandes cantidades de datos [23].

Dentro del campo de la inteligencia artificial se encuentra el Machine Learning, este aprendizaje es un proceso automatizado que extrae patrones de los datos para la construcción de modelos que permiten la predicción utilizando algoritmos supervisados [24], [25]. Una de las modalidades de aprendizaje que tiene Machine Learning es el aprendizaje supervisado, cuya función se basa en entrenar al algoritmo otorgándole las preguntas, llamadas características, y las respuestas, llamadas clases, para que el algoritmo las combine y pueda hacer predicciones [26].

Dentro de las dos técnicas de aprendizaje supervisado, se encuentra la clasificación, los algoritmos de clasificación buscan patrones que luego les permitirán clasificar los elementos y determinar a qué grupos o clases pertenecen. Cabe mencionar que los valores para estos algoritmos deben ser valores discretos [27]. Entre los algoritmos de clasificación se encuentra Kernel, que extiende la regresión logística regular, utilizada para la clasificación binaria, para tratar datos que no son linealmente separables [28].

Dado lo descrito, surge la necesidad de diseñar un modelo predictivo de las actitudes personales y sociales de los estudiantes universitarios, que permita optimizar los servicios que ofrece la institución superior. Teniendo en cuenta aún más que los factores personales y sociales de los estudiantes están relacionados con su rendimiento académico [29], [30]. La investigación cobra aún más relevancia, debido a que, como se indica en [31], la mayoría de las investigaciones de análisis predictivo están relacionadas con la educación primaria y secundaria, por lo que existe un número reducido de aplicaciones en la educación superior que sirven a las instituciones como fuente base para el mejoramiento de la calidad educativa.

En este sentido, el objetivo principal de este artículo es determinar el modelo de aprendizaje supervisado utilizando la técnica Classification Learner para el sistema predictivo de actitudes personales y sociales de estudiantes universitarios de carreras profesionales de ingeniería. Para ello, analizará las métricas de desempeño como Exactitud (A), Precisión (P), Sensibilidad (S) y Especificidad (R), para determinar el algoritmo más adecuado para el modelo predictivo, además, la matriz de confusión y la se identificará la curva característica operativa del receptor (ROC) del modelo.

La investigación tiene como finalidad generar un aporte significativo, para la toma de acciones preventivas y correctivas que permitan cumplir con los estándares de calidad del servicio educativo, cuyo mejoramiento se verá reflejado en la satisfacción de los estudiantes y el rendimiento académico.

La estructura del desarrollo de la investigación se divide en la metodología donde se detalla el nivel de investigación, los participantes, las técnicas de recolección de datos, la validación de

los datos recopilados y el diseño de la metodología de aprendizaje supervisado a través de la técnica de clasificación. A continuación se presentan los resultados y la discusión de los mismos, para finalmente realizar las conclusiones de la investigación.

II. METODOLOGIA DE LA INVESTIGACION

A. Nivel de investigación

La investigación es de carácter descriptivo, ya que se enfoca en determinar el modelo más óptimo de aprendizaje supervisado utilizando la técnica Classification Learner para el sistema predictivo de actitudes personales y sociales de los estudiantes universitarios de las carreras profesionales de ingeniería, a través del análisis de las métricas de desempeño de los algoritmo obtenido (Exactitud (A), Precisión (P), Sensibilidad (S) y Especificidad (R)).

La investigación parte de la identificación de un problema, relacionado con el mejoramiento de la calidad del servicio educativo, el cual se ve reflejado en la satisfacción de los estudiantes universitarios. La satisfacción estudiantil abarca diferentes dimensiones, este estudio se enfoca en la autopercepción de los estudiantes de ingeniería de una universidad pública del Perú con respecto a las actitudes personales y sociales. Para ello, se utilizan métodos o herramientas ya definidas como los sistemas predictivos mediante la modalidad de aprendizaje supervisado.

Esta investigación también busca diseñar un modelo multidimensional predictivo que pueda ser utilizado para crear y almacenar nuevos datos para la institución superior. Con base en esta herramienta tecnológica, determina patrones y calcula reglas de asociación, brindando soporte y confiabilidad a los resultados obtenidos.

B. Participantes

Los participantes en esta investigación están conformados por estudiantes de sexto a décimo ciclo de escuelas profesionales de ingeniería, con un total de 715 estudiantes. Este criterio de selección forma parte de un reglamento establecido y aprobado por la universidad pública del Perú. Debido al carácter obligatorio de la encuesta, fue posible recopilar datos de todos los participantes.

TABLA I. INDICADORES QUE MIDEN LA AUTOPERCEPCIÓN DE ACTITUDES PERSONALES Y SOCIALES

Indicador	es de artículos
1	Asume tus estudios con responsabilidad, seriedad y dedicación
2	Con el orgullo de pertenecer a la universidad
3	Con el compromiso de levantar el nombre de la universidad
4	Con el respeto que muestras a las autoridades, docentes y personal administrativo
5	Con el respeto que tratas a tus compañeros.
6	Con el trato que recibes de tus compañeros
7	Con su interés por ser mejores cada día
8	Con tu compromiso con la sociedad que te rodea

C. Técnica e Instrumento de Recopilación de Datos

La técnica de recolección de datos es la encuesta, y el instrumento utilizado para recolectar datos sobre la autopercepción de los estudiantes de ingeniería sobre las actitudes personales y sociales es el cuestionario, el cual se realizó en línea, a través de la plataforma virtual de la universidad, debido a la implementación de El proceso de enseñanza-aprendizaje en línea en el contexto de la emergencia sanitaria. Además, se dio acceso al cuestionario en línea a través del código de estudiante de cada participante.

Es necesario indicar que el cuestionario estuvo conformado por 8 indicadores (características), los cuales se detallan en la Tabla I. Asimismo, las respuestas al cuestionario para el presente estudio se transformaron en una escala tipo Likert de 4 niveles (1: insatisfecho, 2: poco satisfecho, 3: satisfecho y 4: muy satisfecho), estos niveles de satisfacción, en el presente análisis representan las clases del modelo predictivo de aprendizaje supervisado.

D. Confiabilidad de los Datos Recolectados

La validación de los datos recolectados se realiza mediante el coeficiente alfa de Cronbach a través del software estadístico SPSS. La Tabla II muestra el resultado de confiabilidad, cuyo valor es 0.962.

E. Diseño de procesamiento de datos

La metodología del proceso de investigación inicia con la recolección de datos a través del cuestionario en línea de la autopercepción de los estudiantes de ingeniería sobre las actitudes personales y sociales. Las respuestas se almacenan en una base de datos, cuya información es procesada por el controlador Open Data Base Connectivity (ODBC) y el software Matlab R2021a.

De esta forma, se inicia el proceso de aprendizaje supervisado, mediante la técnica de Classification Learner, se identifica el mejor algoritmo de Machine Learning, mediante la evaluación de métricas de desempeño, luego de lo cual se realiza la predicción, que permitirá el análisis de los resultados. En la Fig. 1 se muestra el esquema de la metodología de aprendizaje supervisor utilizado.

TABLA II. PRUEBA ALFA DE CRONBACH

Estadísticas de confiabilidad	
Alfa de Cronbach	Nº de elementos
0.962	8

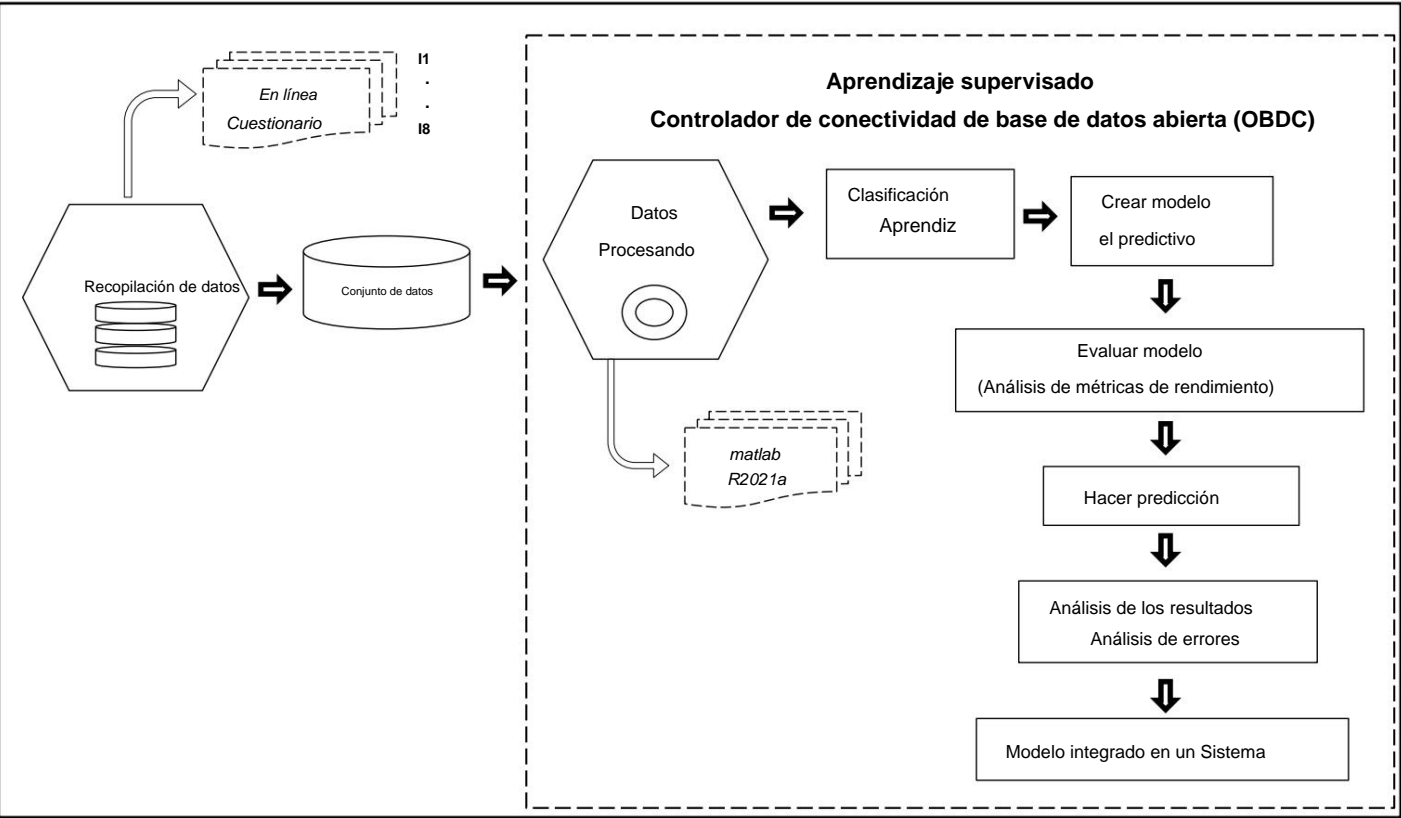


Figura 1. Metodología de Aprendizaje Supervisado.

tercero RESULTADO Y DISCUSIÓN

A. Determinación del Modelo Predictivo

Por medio del software Matlab R2021a, y a través de la técnica Classification Learner de Machine Learning Toolbox 12.1, se identifica el mejor modelo predictivo determinado por la validación de la precisión, en términos estadísticos se relaciona con el sesgo de una estimación y se representa como la proporción de resultados verdaderos (verdaderos positivos y verdaderos negativos) dividida por el número total de casos examinados (verdaderos positivos, falsos positivos, verdaderos negativos, falsos negativos).

En la Fig. 2 se muestran los resultados generados, los cuales muestran que el algoritmo Kernel: Logistic Regression Kernel, es el que presenta una mejor precisión del 86.9% para el sistema predictivo de las actitudes personales y sociales de los estudiantes de ingeniería. Le sigue los algoritmos Tree: Coarse Tree con una precisión del 86,6 %, SVM: Coarse Gaussian SVM con una precisión del 86,4 % y SVM: Linear SVM con una precisión del 86,2 %.

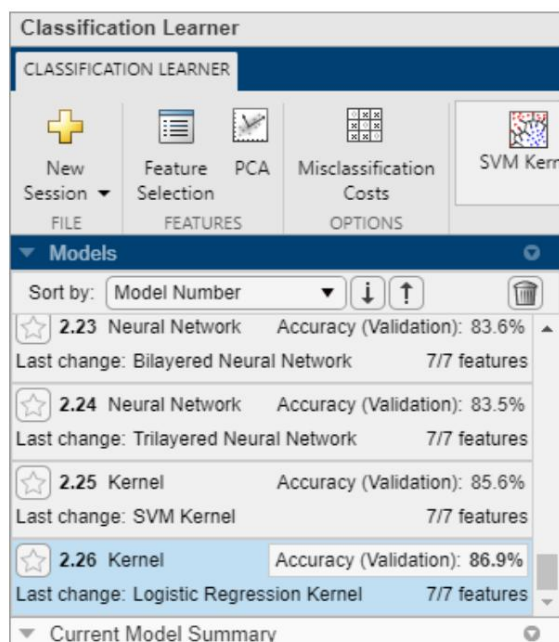


Fig. 2. Determinación del Algoritmo de Clasificación.

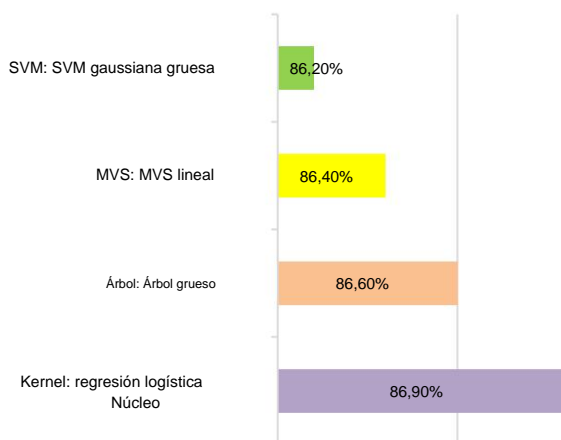


Fig. 3. Análisis comparativo de los resultados de precisión.

Asimismo, en la figura 3 se muestran los cuatro algoritmos con mayor precisión, lo que identificó la técnica de clasificación, donde se observa como mejor el algoritmo Logistic Regression Kernel.

Siendo la precisión el parámetro que mide el porcentaje de casos que ha acertado el modelo, según la Fig. 3, se puede decir que el modelo predictivo mediante el algoritmo Logistic Regression Kernel ofrece un total del 86,9% del número de predicciones positivas que será correcto, es decir, el valor se refiere a qué tan cerca está el resultado de una medición del valor verdadero.

B. Resultados de las Métricas del Modelo Predictivo

Como parte del análisis de las métricas de desempeño, visualizaremos la matriz de confusión del algoritmo Logistic Regression Kernel, la matriz de confusión nos permite visualizar el desempeño de un algoritmo de aprendizaje supervisado y cada columna de la matriz representará el número de predicciones de cada clase (1: insatisfecho, 2: poco satisfecho, 3: satisfecho y 4: muy satisfecho), mientras que cada fila representa las instancias de la clase real, es decir este análisis nos permite ver qué tipos de aciertos y errores nuestro sistema predictivo tiene .

En la Fig. 4, se muestra la matriz de confusión, donde se muestra el porcentaje de prueba de falsos negativos (FNR), también llamada tasa de error, es la probabilidad de que un verdadero positivo no lo haga, también se muestra la tasa de verdaderos positivos (TPR), que mide la métrica de sensibilidad, esta métrica viene a hacer la probabilidad de que un resultado positivo real sea positivo.

Como se puede observar en la Fig. 4, de las 4 clases sobre las que actúa el modelo predictivo, la clase 3 (satisfecho) muestra un 89,9% de sensibilidad y un 10,2% de falsos negativos, esto quiere decir que el modelo predictivo tiene la capacidad de discriminar entre un verdadero positivo (TP) de un falso negativo (FN), es decir, hay un 89,9% de capacidad para poder detectar correctamente alumnos satisfechos entre alumnos insatisfechos. Por otro lado, la clase 2 (no muy satisfecho) muestra un menor porcentaje de sensibilidad igual al 64,4% con un porcentaje de falsos negativos del 35,6%, es decir tiene la capacidad de detectar solo el 35,6% de alumnos satisfechos entre los alumnos insatisfechos. .

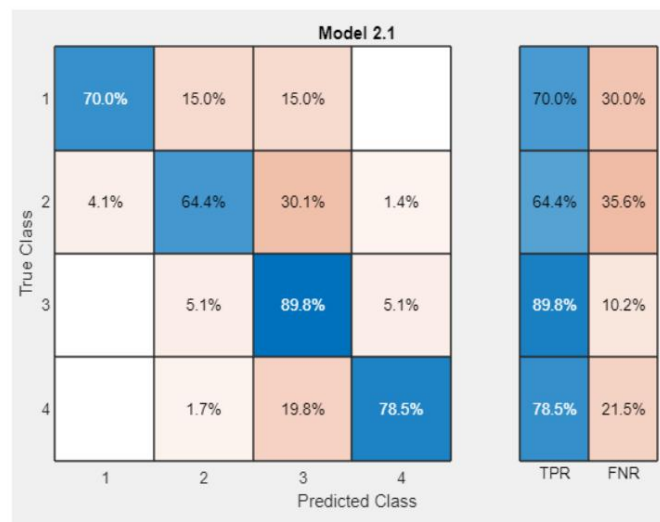


Figura 4. Matriz de confusión basada en las tasas de TPR y FNR.

En la figura 5 se muestra la matriz de confusión, donde se muestran el valor predictivo positivo (PPV) y la tasa de falsos descubrimientos (FDR). Cabe señalar que la métrica de precisión mide la calidad del modelo de aprendizaje automático en tareas de clasificación, se debe tener en cuenta que a menor valor de dispersión, mayor precisión del modelo.

Como se puede observar en la Fig. 5, el modelo predictivo para la clase 3 (satisfecho) presenta el mayor valor de precisión, en este caso es del 87,3% y un porcentaje de FDR del 12,7%, es decir, solo el 87,3% de los alumnos estarán realmente satisfechos. satisfecho con las actitudes personales y sociales percibidas, mientras que el 12,7% de los ejemplos se equivocará en la predicción.

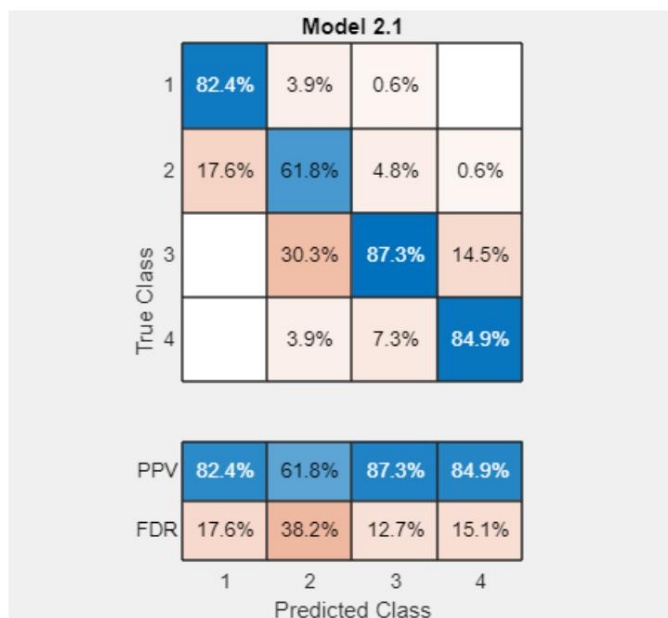


Figura 5. Matriz de confusión basada en tasas de PPV y FDR.

■ Sensibilidad ■ especificidad ■ Exactitud ■ Precisión

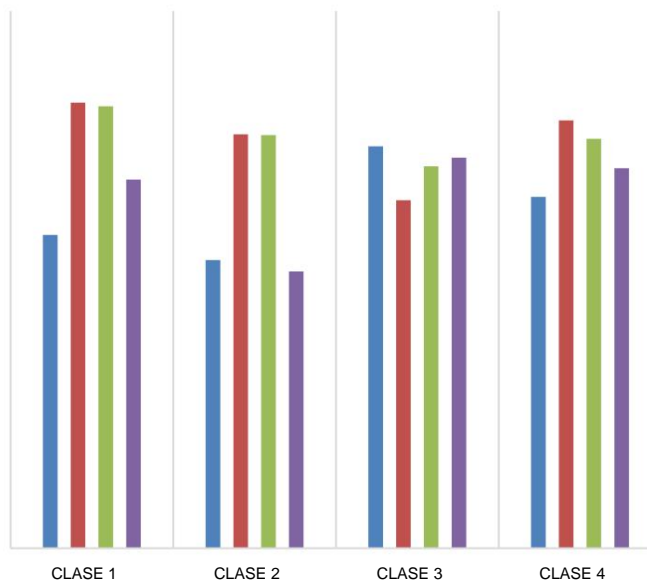


Fig. 6. Métricas de las Cuatro Clases del Sistema Predictivo.

En la Fig. 6 se muestran las métricas del modelo predictivo de las cuatro clases (1: insatisfecho, 2: poco satisfecho, 3: satisfecho y 4: muy satisfecho) mediante el algoritmo Logistic Regression Kernel, con esto se puede decir, en general, la Precisión es 79,09%, la Sensibilidad es 75,66%, la Especificidad es 92,09% y la Exactitud es 91,96%.

A continuación, se realizará el análisis de las curvas de características operativas del receptor (ROC) del algoritmo Logistic Regression Kernel, el cual constituye un método estadístico para determinar la precisión del modelo, esta prueba se realiza con tres propósitos, determinar el corte -punto de partida de una escala continua en la que se alcance la mayor sensibilidad y especificidad, evaluar la capacidad de diferenciar alumnos satisfechos e insatisfechos, y comparar la capacidad discriminativa de dos o más diagnósticos que expresen sus resultados como escalas continuas. Es necesario especificar cuanto más se acerque a 1 el valor del área sobre la curva (AUC), y el modelo tendrá un mejor desempeño y mayor precisión.

En ese sentido, en la Fig. 7, el gráfico ROC para la clase 1 (insatisfecho) se muestra.

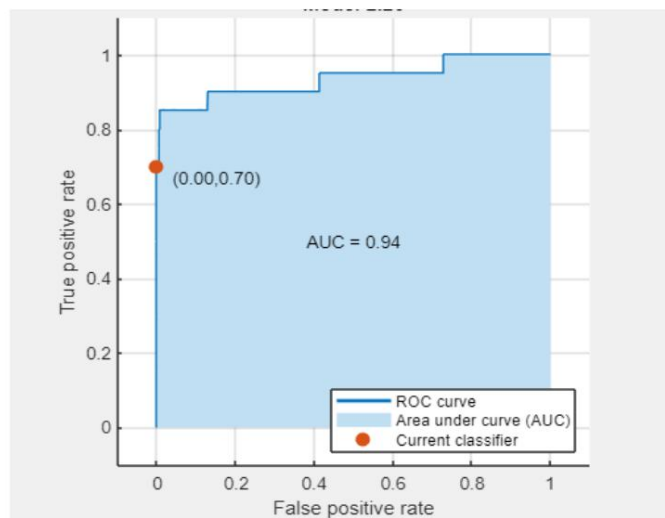


Fig. 7. Gráficos ROC para Clase 1.

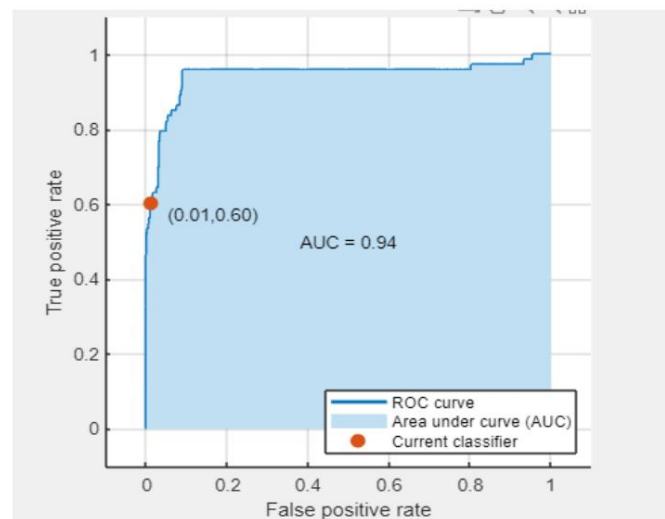


Fig. 8. Gráficos ROC para Clase 2.

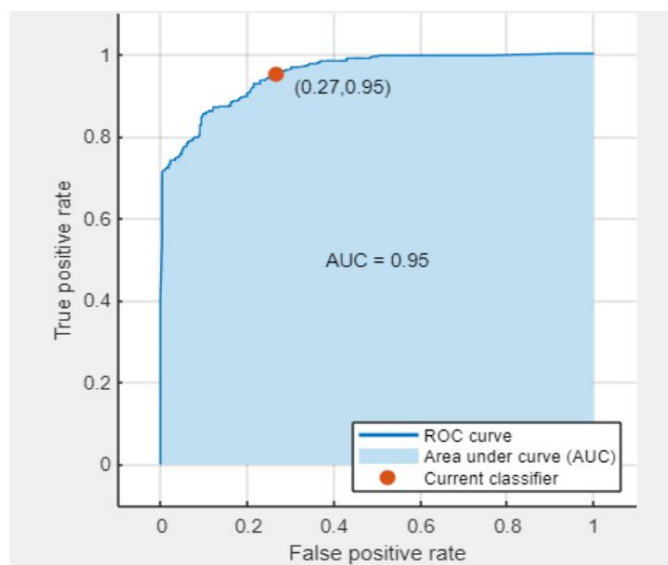


Fig. 9. Gráficos ROC para Clase 3.

En la Fig. 7 se evidencia que para la clase 1 existe una sensibilidad del 94%. Además, el umbral de discriminación es de 0,70 para la tasa de verdaderos positivos y de 0,00 para la tasa de falsos positivos, mostrando un valor óptimo del área en la curva (AUC) de 0,93.

En la Fig. 8 se muestra el gráfico ROC para la clase 2 (no muy satisfecha).

En la Fig. 7 se evidencia que para la clase 1 existe una sensibilidad del 94%. Además, el umbral de discriminación es de 0,70 para la tasa de verdaderos positivos y de 0,00 para la tasa de falsos positivos, mostrando un valor óptimo del área en la curva (AUC) de 0,93.

En la Fig. 8 se muestra el gráfico ROC para la clase 2 (no muy satisfecha).

En la Fig. 9 se evidencia que para la clase 3 existe una sensibilidad del 95%. Además, el umbral de discriminación es de 0,95 para la tasa de verdaderos positivos y de 0,27 para la tasa de falsos positivos, mostrando un valor óptimo del área en la curva (AUC) de 0,95.

En ese sentido, en la Fig. 10 se muestra el gráfico ROC para la clase 4 (muy satisfecha). Para la clase 4 se evidencia una sensibilidad del 97%. Además, el umbral de discriminación es de 0,78 para la tasa de verdaderos positivos y de 0,03 para la tasa de falsos positivos, mostrando un valor óptimo del área en la curva (AUC) de 0,97.

Finalmente, se muestra en la Fig. 11, la gráfica de coordenadas paralelas, utilizada para graficar datos multivariados, esta gráfica muestra la relación entre los indicadores de autopercepción de los estudiantes de ingeniería sobre las actitudes personales y sociales entre las cuatro clases del modelo predictivo (1: insatisfecho, 2: poco satisfecho, 3: satisfecho y 4: muy satisfecho).

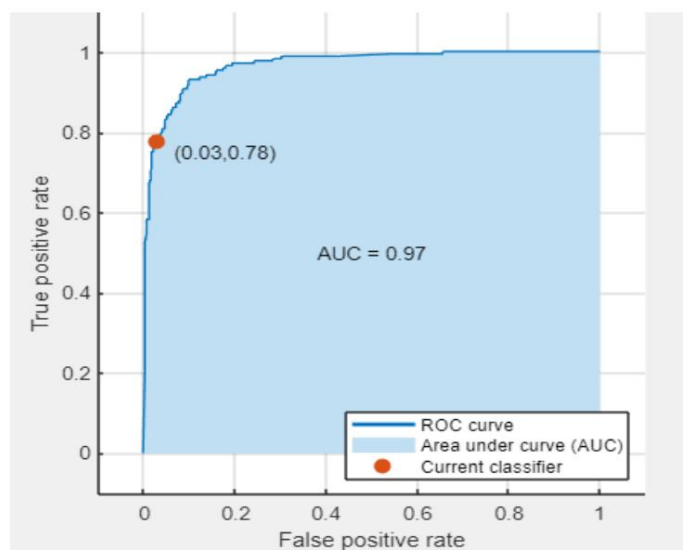


Fig. 10. Gráficos ROC para Clase 4.

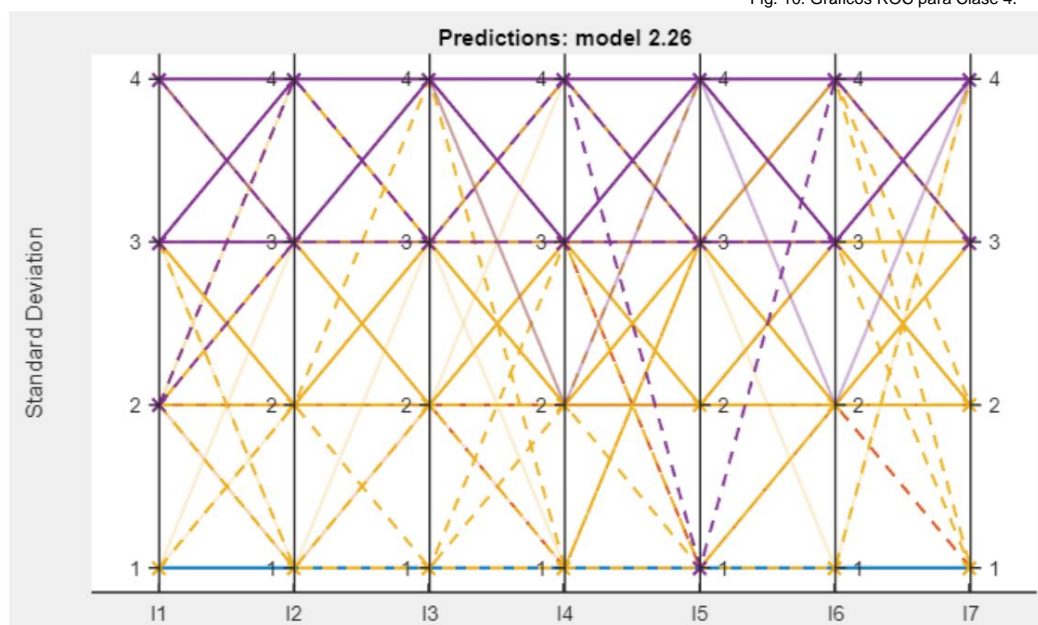


Fig. 11. Gráfico de coordenadas paralelas.

Cabe señalar que los datos están vinculados a 8 descriptores relacionados con los niveles de satisfacción de la autopercepción de los estudiantes de ingeniería sobre las actitudes personales y sociales, durante el proceso de enseñanza-aprendizaje virtual, dichos datos son de tipo cualitativo ordinal, definiendo así el cuatro clases, así mismo, como parte de la determinación del modelo predictivo utilizando el software Matlab R2021a, se utilizó la información recolectada en base a los primeros 7 indicadores, que se denominan predictores, y el indicador de la variable bajo análisis (que representa la satisfacción de la autopercepción de los estudiantes de ingeniería sobre las actitudes personales y sociales), está definida por el indicador final (I8), por ser el que generalmente engloba a la variable bajo análisis, por lo que en la Fig. 11 solo se muestra a partir del indicador I1 al indicador I7.

C. Discusión

Los resultados de las métricas del modelo de predicción a través del algoritmo Logistic Regression Kernel, en relación a la satisfacción con las actitudes personales y sociales, presentan en general, en sus 4 clases, una precisión de 79.09%, una Sensibilidad de 75.66%, una Especificidad de 92.09% y una Precisión del 91,96%.

También el gráfico ROC para la clase 4 (muy satisfecho) muestra una sensibilidad del 97%, un umbral de discriminación es de 0,78 para la tasa de verdaderos positivos y de 0,03 para la tasa de falsos positivos, mostrando un valor óptimo del área en la curva (AUC) de 0.97.

En cuanto al valor óptimo de las métricas de rendimiento se sostiene que son valores óptimos, con base en la investigación de [18] donde se afirma que se realiza un modelo para predecir el rendimiento académico de los estudiantes entrantes a través del aprendizaje automático, los resultados arrojaron que es factible predecir el rendimiento, ya que el modelo tiene un 69% de precisión. De igual forma, en [26] se indica que los resultados muestran una precisión del 82%, por lo tanto, se puede señalar que el modelo predictivo tendrá un desempeño óptimo al ser implementado.

En cuanto a la ROC obtenida de 97% en la presente investigación, en [20] el autor señala que su modelo predictivo fue bueno ya que su precisión general fue de 75.42% y un área bajo la curva ROC de 0.805. Esto se basa en la teoría establecida de que cuanto más cerca esté este valor de 1, mejor funcionará el modelo y más preciso será. De igual forma en [23], el autor señala que un AUC del 60% del 91% o 99%, representa un mejor desempeño del algoritmo clasificador, siendo estos resultados favorables para la investigación.

En [29] se señala que en relación a la precisión del modelo para la clase de rendimiento académico, el valor obtenido fue 62.45% inferior al valor obtenido en la investigación realizada por [24] quienes en su modelo de predicción, obtuvo el valor 73%, muy cercano al resultado obtenido por [25] que fue 69%.

Comparando los resultados con la investigación realizada por [31] donde se obtuvo un valor de precisión menor del 80% en el sistema predictivo para los datos de entrenamiento y del 76% para los datos de validez, este estudio presenta resultados de rendimiento superiores. A su vez, en la investigación de [27] se obtuvo un 82,87% de precisión utilizando el algoritmo del árbol de decisión, lo que representa un valor inferior al resultado de nuestra investigación.

En el estudio realizado por [19] se utilizaron datos de 914 estudiantes para predecir su clasificación final (aprobado o reprobado), con un modelo predictivo de aprendizaje supervisado y la técnica Classification Learner y así obtener un modelo para predecir los resultados de la estudiantes, mostrando que con el algoritmo de red (Naive de Bayes) que mostró óptimo precisión con un 71,0%, en comparación con las cuatro técnicas analizadas (redes neuronales, regresión logística, árbol de decisión y red bayesiana) con mayor porcentaje para la clase aprobada y menor para suspenso.

IV. CONCLUSIÓN

Con los resultados obtenidos se evidencia que es posible aplicar la metodología de aprendizaje supervisado utiliza técnicas de Clasificación del Aprendizaje para el sistema predictivo de las actitudes personales y sociales de los estudiantes, a través de la gráfica de coordenadas paralelas la asociación y/o relación de los indicadores de la variable bajo análisis con las cuatro clases del modelo predictivo a través del algoritmo Logistic Regression Kernel. Se concluyen los resultados de las métricas del modelo de predicción en relación a la satisfacción con las actitudes personales y sociales, en general presenta un desempeño óptimo de sus métricas de validación en sus 4 clases, con una precisión de 79.09%, una Sensibilidad de 75.66%, una Especificidad de 92,09% y una Precisión del 91,96%.

Se recomienda ampliar la línea de investigación a otras con otros indicadores relacionados con la satisfacción de los estudiantes con el servicio educativo, porque no solo permite a la institución superior tener una base de datos o referencia a través de una técnica de clasificación rápida y confiable para tomar medidas preventivas y correctivas acciones, para mejorar la calidad de la educación, pero también el sistema predictivo influye en la reducción de las tasas de deserción y mejorar el rendimiento académico de los estudiantes.

RECONOCIMIENTO

Agradecemos a los investigadores que han aportado su conocimiento en el desarrollo de este trabajo.

REFERENCIAS

- [1] FE Ceballos, JE Rojas, LG Cuba, LP Medina y AR Velazco, "Análisis de la calidad de los servicios en los centros universitarios", *Universidad, Ciencia y Tecnología*, vol. 25, núm. 108, págs. 23-29, 2021. DOI: <https://doi.org/10.47460/uct.v25i108.427>.
- [2] H. Aldowah, H. Al-Samarraie y WM Fauzy, "Extracción de datos educativos y análisis del aprendizaje para la educación superior del siglo XXI: revisión y síntesis", *Telemática e informática*, vol. 37, págs. 13-49, 2019. DOI: <https://doi.org/10.1016/j.tele.2019.01.007>.
- [3] D. Vlachopoulos, "Enseñanza de calidad en la educación superior en línea: las perspectivas de 250 tutores en línea sobre tecnología y pedagogía" *Revista Internacional de Tecnologías Emergentes en el Aprendizaje (IJET)*, vol. 16, DOI: <https://doi.org/10.3991/ijet.v16i06.201740-56>, 2021.
- [4] B. Bahati, U. Fors, P. Hansen, J. Nouri y E. Mukama, "Medir la satisfacción del alumno con las estrategias de evaluación electrónica formativa" *Revista Internacional de Tecnologías Emergentes en el Aprendizaje (IJET)*, vol. 14, No. 7, págs. 246-247, 2019. DOI: <https://doi.org/10.3991/ijet.v14i07.9120>.
- [5] B. Semprún, y K. Ferrer, "Satisfacción de los estudiantes de un curso de Bioquímica: una evaluación después de aplicar estrategias neurodidácticas", *Revista San Gregorio*, vol. 1, no. 38, págs. 1-14, 2020. DOI: <https://dx.doi.org/rsan.v1i38.1241>.
- [6] I. Amoako y K. Asamoah-Gyimah, "Indicadores de la satisfacción de los estudiantes con los servicios de educación de calidad en algunas universidades seleccionadas en

- Ghana", *South African Journal of Higher Education*, vol. 35, núm. 4, págs. 61-72, 2020. DOI: 10.20853/34-5-4252.
- [7] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang y G. Cong, "Análisis de la satisfacción del usuario con las plataformas de educación en línea en China durante la pandemia de COVID-19", *Healthc.*, vol. 8, núm. 3, pág. 200, 2020. DOI: 10.3390/healthcare8030200.
- [8] E. Alvarado-Lagunas, J. Luyando-Cuevas y E. Piccaso-Palencia, "Percepción de los estudiantes hacia la calidad de las universidades privadas en Monterrey", *Revista Iberoamericana de Educación Superior*, vol. 6, núm. 17, págs. 58-76, 2015. DOI: 10.1016/j.rides.2015.10.003.
- [9] K. Mukhtar, K. Javed, M. Arooj y A. Sethi, "Ventajas, limitaciones y recomendaciones para el aprendizaje en línea durante la era de la pandemia de COVID-19", *Pak. J. Med. ciencia*, vol. 36, núm. COVID19-S4, págs. S27-S31, 2020, doi: 10.12669/pjms.36.COVID19-S4.2785.
- [10] W. Niebles-Núñez y M. Cabarcas-Velásquez, "Responsabilidad social: elemento formativo de los estudiantes universitarios", *Revista Latinoamericana de Estudios Educativos*, vol. 14, núm. 1, págs. 257-268, 2018. DOI: https://doi.org/10.17151/rlee.2018.14.1.6.
- [11] A. Bernasconi y E. Rodríguez-Ponce, "Análisis exploratorio de las percepciones sobre los estilos de liderazgo, el clima académico y la calidad de la educación de pregrado", *Formulario. Universidad*, vol. 11, núm. 3, págs. 29-40, 2018, DOI: http://dx.doi.org/10.4067/S0718-50062018000300029.
- [12] P. Ramkissoon, L.J. Belle y T. Bhurosy, "Percepciones y experiencias de los estudiantes sobre el uso de tecnologías interactivas de aprendizaje en línea en Mauricio", *Revista Internacional de Evaluación e Investigación en Educación*, vol. 9, núm. 4, págs. 833-839, 2020. DOI: http://doi.org/10.11591/ijere.v9i4.20692.
- [13] A. Ali y F. Mohammed, "Medición de la calidad de E-Learning y Desaire2Learn en la Facultad de Ciencias y Humanidades de Alghat, Universidad de Majma-ah", *Revista Internacional de Ciencias Informáticas Avanzadas y Aplicaciones (IJACSA)*, vol. 9, núm. 1, págs. 61-72, 2020. DOI: 10.14569/IJACSA.2018.090170.
- [14] H. Hafiza y R. Ibrahim, "Educación a distancia durante la pandemia de COVID-19: las percepciones y preferencias de los estudiantes universitarios en Malasia hacia el aprendizaje en línea", *Revista internacional de informática avanzada y aplicaciones (IJACSA)*, vol. 12, núm. 4, págs. 118-126, 2021. DOI: 10.14569 / IJACSA.2021.0120416.
- [15] C. Coman, LG jiru, L. Meseñan-Schmitz, C. Stanciu y MC Bularca, "Enseñanza y aprendizaje en línea en la educación superior durante la pandemia del coronavirus: la perspectiva de los estudiantes", *Sostenibilidad*, vol. 12, núm. 24, págs. 1-22, 2020. DOI: 10.3390/su122410367.
- [16] F. Phang, et al., "Integrating Drone Technology in Service Learning for Engineering Students", *International Journal of Emerging Technologies in Learning (IJET)*, vol. 16, núm. 15, págs. 78-96, 2021. DOI: https://doi.org/10.3991/ijet.v16i15.23673.
- [17] A. Ramos, M. Aldude, J. Estrada, V. Señas y L. Andrade-Arenas, "Análisis del uso de herramientas tecnológicas en la educación superior universitaria utilizando la metodología de sistemas blandos", *International Journal of Advanced Computer Science y Aplicaciones (IJACSA)*, vol. 11, núm. 7, págs. 412-420, 2020, DOI: 10.14569 / IJACSA.2020.0110754.
- [18] H. Sadiq, A. Dahan, Neama, FM Ba-Alwi y N. Ribata, "Extracción de datos educativos y análisis del rendimiento académico de los estudiantes mediante weka", *Revista indonesia de ingeniería eléctrica e informática*, vol. 9, núm. 2, págs. 447-459, 2018. DOI: https://doi.org/10.11591/ijeecs.v9i2.pp447-459.
- [19] CH Menacho, "Predicción del rendimiento académico aplicando técnicas de minería de datos", *Learning and Individual Scientific Annals*, vol. 78, núm. 1, págs. 26-33, 2017. DOI: http://dx.doi.org/10.21704/ac.v78i1.811.
- [20] RS Baker y K.Yacef, "El estado de la minería de datos educativos en 2009: una revisión y visiones futuras", *Revista de minería de datos educativos*, vol. 1, No. 1, págs. 3-16. DOI: https://doi.org/http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314.
- [21] EB Costa, B. Fonseca, MA Santana, FF De Araújo y J. Rego, "Evaluación de la efectividad de las técnicas educativas de minería de datos para la predicción temprana del fracaso académico de los estudiantes en los cursos de introducción a la programación", *Computers in Human Behavior*, vol. 73, págs. 247-256, 2017. DOI: https://doi.org/10.1016/J.CHB.2017.01.047.
- [22] B. Bakhshinategh, OR Zaiane, S. ElAtia y D. Ipperciel, "Aplicaciones y tareas de minería de datos educativos: un estudio de los últimos 10 años", *Educación y tecnologías de la información*, vol. 23, núm. 1, págs. 537-553. DOI: https://doi.org/10.1007/s10639-017-9616-z.
- [23] D. Moonsamy, N. Naicker, TT Adeliyi y RE Ogunsakin, "Un metaanálisis de la minería de datos educativos para predecir el rendimiento de los estudiantes en la programación", *Revista internacional de informática avanzada y aplicaciones (IJACSA)*, vol. 12, núm. 2, págs. 97-104, 2021. DOI: 10.14569 / IJACSA.2021.0120213.
- [24] OD Castrillón, W. Sarache, William y S. Ruiz-Herrera, "Predicción del rendimiento académico mediante técnicas de inteligencia artificial", *Educación universitaria*, vol. 13, núm. 1, págs. 93-102, 2020. DOI: https://doi.org/10.4067/S0718-50062020000100093.
- [25] D. Buenaño-Fernández, D. Gil y S. Luján-Mora, "Aplicación del aprendizaje automático en la predicción del rendimiento para estudiantes de ingeniería informática: un estudio de caso", *Sostenibilidad*, vol. 11, núm. 10, pág. 2833. DOI: https://doi.org/10.3390/su11102833.
- [26] R. Katarya, J. Gaba, A. Garg, Aryan y V. Verma, "Una revisión de los sistemas de predicción del rendimiento académico de los estudiantes basados en el aprendizaje automático", *Conferencia internacional sobre inteligencia artificial y sistemas inteligentes (ICAIS) de 2021*, págs. 254 -259, Coimbatore, India: IEEE, 2021. DOI: https://doi.org/10.1109/ICAIS50930.2021.9395767.
- [27] P. Sokkhey y T. Okazaki, "Estudio sobre el factor dominante para la predicción del rendimiento académico utilizando métodos de selección de características", *Revista internacional de informática avanzada y aplicaciones (IJACSA)*, vol. 11, núm. 8, págs. 492-502, 2020. DOI: 10.14569/IJACSA.2020.0110862.
- [28] F. Makombe y M. Lall, "Un modelo predictivo para la determinación del rendimiento académico en instituciones privadas de educación superior", *Revista Internacional de Aplicaciones y Ciencias Informáticas Avanzadas (IJACSA)*, vol. 11, 9, págs. 415-419, 2020. DOI: 10.14569/IJACSA.2020.0110949.
- [29] B. Díaz-Landa, R. Meleán-Romero y W. Marín-Rodríguez, "Rendimiento académico de los estudiantes de Educación Superior: predicciones de los factores que influyen en la decisión de los árboles", *Redalyc*, vol. 23, núm. 3, págs. 616-635, 2021. DOI: https://doi.org/10.36390/telos233.08.
- [30] M. Bravo, S. Salvo y C. Muñoz, "Perfiles de estudiantes chilenos según rendimiento académico en matemáticas: un estudio exploratorio utilizando árboles de clasificación y bosques aleatorios", *Estudios en Evaluación Educativa*, vol. 44, págs. 50-59, 2015. DOI: https://doi.org/10.1016/j.stueduc.2015.01.002.
- [31] M. Imran, S. Latif, D. Mehmood y MS Shah, "Predicción del rendimiento académico de los estudiantes mediante técnicas de aprendizaje supervisado", *Revista Internacional de Tecnologías Emergentes en el Aprendizaje (IJET)*, vol. 14, No. 14, 92-104, 2019. DOI: https://doi.org/10.3991/ijet.v14i14.10310.