

Algoritmo de aprendizaje automático y puntuación de factores para predecir la consejería estudiantil

Nusrat Jahan, Saiful Islam, Rezwana Sultana

Resumen: La realización personal es una de las mejores cosas para una vida exitosa. A veces, uno necesita ayuda para darse cuenta de los malos hábitos, las metas profesionales y lograr la salud mental, así como para superar otros problemas. Esta ayuda se conoce generalmente como "Consejería". Para garantizar la eficacia del servicio de asesoramiento, la principal preocupación es averiguar el grupo objetivo de instancias. Muchos investigadores trabajaron con la predicción del rendimiento de los estudiantes en función de los atributos académicos; además, también se necesita asesoramiento de los estudiantes para aumentar su rendimiento. Hemos abordado este problema para este papeleo. Aquí, se propone un modelo para predecir un estudiante que necesita orientación. El primero fue investigar a estudiantes universitarios que sienten la urgencia de recibir consejería de ayuda psicológica a partir de sus circunstancias y el segundo fue predecir de manera eficiente qué grupo de estudiantes realmente necesita consejería. Este papeleo se estableció con 498 instancias y cada una constaba de 6 atributos.

En el caso de evaluar el resultado, el papel muestra superioridad sobre los métodos más avanzados para predecir la orientación de los estudiantes a través del aprendizaje automático y el método de puntuación de factores. Aplicamos una validación cruzada de 10 pliegues y un método de evaluación de divisiones de conjuntos de datos del 66% para encontrar un mejor algoritmo entre los 5 algoritmos seleccionados que son Ibk, Naive Bayes, Multilayer, SMO y Random Forest. Weka 3.8.0 se ha utilizado para algoritmos de aprendizaje automático donde Ibk (aprendizaje basado en instancias) resultó ser el mejor para nuestro enfoque con una precisión del 95,38 %.

Palabras clave: Consejería Estudiantil, Factor Scoring, Máquina Aprendizaje, Ibk, Weka

I. INTRODUCCIÓN

En todo el mundo la educación es una de las necesidades básicas.

Sin embargo, la ruta de la educación a veces crea un problema agudo entre los estudiantes. Según el cuartel general de la policía en Bangladesh, más de 10,000 personas se suicidan cada año por diversas razones, donde, el Dr. Mohammad Jahangirul Alam, médico de la Universidad Médica Bangabandhu Sheikh Mujib (BSMMU), Bangladesh, dijo que "El trastorno subyacente más común de un ser humano es depresión y 30-70

por ciento de las víctimas de suicidio sufren de depresión mayor o trastorno bipolar (maníaco depresivo)" [1]. Aunque los estudiantes son muy ambiciosos acerca de su academia y su carrera, aún así rara vez tienen ideas de cómo lograr sus ambiciones también. Una proporción muy pequeña de los estudiantes completó su graduación de acuerdo con su plan. Al mismo tiempo, un gran número de

los estudiantes ingresan a una universidad con un objetivo profesional específico, pero desde el comienzo de su vida universitaria se postergan de muchas maneras, por ejemplo, redes sociales, mala influencia, adicción a las drogas, desmotivación para estudiar, etc. La mayor parte del tiempo en nuestro país, Los padres ocasionalmente consideran el interés de sus hijos sobre la educación, como resultado crea distancia entre padres e hijos y tiene efectos negativos a largo plazo en su futuro. Como estudiante sufren por la

resto de su vida, incluso si eligen una manera de convertirse exitosos y debido a la frustración algunos de ellos pueden volverse desmotivados. Para minimizar este dolor, la motivación y el asesoramiento específicos del sujeto planificado previamente pueden llevarlos por el camino correcto en el momento adecuado. En todo el mundo, se ha convertido en una preocupación primordial mejorar el desarrollo personal de un estudiante. Para garantizar el desarrollo mundial de un país, nutrir a los estudiantes en función de su elección de carrera es la máxima prioridad cada día que pasa. En nuestro país, muchas universidades brindan zona de consejería estudiantil pero la calidad y el proceso de esta instalación aún no es considerable. Talia Deniozou, junio de 2015 realizó una encuesta donde participaron 1189 estudiantes universitarios. Sin embargo, 993 de ellos eran de pregrado y alrededor del 84% de ellos respondieron correctamente. Aquí, una pregunta era para estudiantes universitarios autoevaluados en función de sus problemas generales de salud mental. El resultado fue que el 35% de los estudiantes estaban mentalmente muy bien, el 24% estudiantes buenos, el 20% se sintieron excelentes, el 16% regulares y finalmente el 5% mentalmente pobres [2].

"Consejería" es ahora una palabra considerable. Encontramos diferentes algoritmos para predecir el estado de consejería (carrera) y otras tareas relacionadas. La mayor parte de la investigación se centró en el análisis del rendimiento académico de los estudiantes en lugar de en los problemas de asesoramiento de los estudiantes basados en el rendimiento académico. Nikita Gorad, Ishani Zalte, Aishwarya Nandi y Deepali Nayak exploraron el uso eficiente de la minería de datos en la carrera. asesoramiento. En su artículo, consideraron Introvertido vs. Extrovertido, Sensitivo vs. Intuitivo, Pensamiento vs. Sentimiento, y Juicio vs. Percepción como factores para predecir la orientación profesional de un estudiante que necesita ayuda para seleccionar la carrera. Según la encuesta realizada por el Consejo de Investigaciones Científicas e Industriales (CSIR) y el hallazgo fue que alrededor del 40% de los estudiantes están confundidos acerca de la planificación de su carrera [3]. Mientras tanto, Pooja Thakar, Anil Mehta y Manisha discutieron sobre la extracción de datos educativos que se basó en diferentes resultados de encuestas, como encuestas sobre estudiantes débiles, determinación de la satisfacción de los estudiantes por un camino en particular, Evaluación de la facultad, Evaluación integral de los estudiantes, Selección de idioma de enseñanza en el aula, Predicción abandono de estudiantes, planificación de matrícula de cursos, evaluación de actividades colaborativas, etc. [4]. Modo y FN definieron la consejería como una relación de ayuda entre el consejero y el aconsejado.

Manuscrito revisado recibido el 27 de septiembre de 2019.

* Autor de correspondencia

Nusrat Jahan*, Informática e Ingeniería, Universidad Internacional Daffodil, Dhaka, Bangladesh. Correo

electrónico: nusratjahan.cse@diu.edu.bd

Saiful Islam, Informática e Ingeniería, Daffodil International University, Dhaka, Bangladesh. Correo electrónico: saiful.cse@diu.edu.bd

Rezwana Sultana, Informática e Ingeniería, Universidad Daffodil, Dhaka, Bangladesh. Correo electrónico: sultana.cse@diu.edu.bd

Algoritmo de aprendizaje automático y puntuación de factores para predecir la consejería estudiantil

La consejería es una forma de ayudar al aconsejado a adaptarse adecuadamente a la familia, la escuela, la asociación de pares y la sociedad también [5]. El sitio de redes sociales es una de las plataformas más grandes donde los estudiantes pasan su tiempo máximo en lugar de su desarrollo personal o estudio, como resultado, afecta su rendimiento académico. Sin embargo, las redes sociales pueden ayudarnos si la forma de uso se centra en la habilidad personal.

desarrollo [6]. Shruthi P. y Chaitra BP, exploraron la predicción del rendimiento académico de los estudiantes en el próximo semestre. Este enfoque también es útil para encontrar estudiantes débiles además de proporcionar una guía de estudio adecuada. En este artículo, utilizaron técnicas de minería de datos para predecir la

rendimiento y descubrió que el algoritmo bayesiano ingenuo proporcionaba la mayor precisión [7]. En esta era reciente, muchos investigadores también hablaron sobre el e-consejería. Emmanuel AK y et al. exploró el sistema de asesoramiento electrónico que se basaba en los datos históricos de los estudiantes. En este artículo, el autor demostró la plataforma de asesoramiento electrónico para estudiantes de secundaria. Su El estudio también se centró en los usos de las TIC para proporcionar servicios de asesoramiento en línea a los estudiantes [8]. Zamani y Shiller identificaron que el correo electrónico, el chat, las videoconferencias y los mensajes de texto, así como los sistemas de mensajes cortos (SMS), son herramientas de asesoramiento electrónico ampliamente utilizadas [9]. N. Banu Priya y I. Juvanna también habló sobre una aplicación de Android que fue implementado para la consejería universitaria en línea. Aquí, ellos proporcionó una plataforma de asesoramiento electrónico entre los estudiantes y un personal de asesoramiento, donde el éxito académico y personal, los inconvenientes, los problemas y los comentarios de los estudiantes fueron el enfoque principal para el asesoramiento [10].

Por el contrario, en la era reciente, el aprendizaje automático es una técnica que nos ayuda a resolver diferentes modos de problemas mediante el conocimiento de la máquina. Hay muchos algoritmos y algunos de ellos se probaron como el mejor algoritmo para varios propósitos.

Los algoritmos de aprendizaje automático se dividen en muchas categorías según su propósito [11]. Árbol de decisión (Bosque aleatorio), Máquina de vectores de soporte, Naive bayes, Neural

Se aplicó el aprendizaje basado en redes e instancias para clasificar nuestro conjunto de datos mediante el uso de la herramienta de aprendizaje automático Weka (versión 3.8.0). Todos estos son algoritmos de aprendizaje supervisado.

Decision Tree es uno de los enfoques de aprendizaje automático supervisado en el que los datos se clasifican continuamente según un determinado parámetro. El árbol se puede explicar con la ayuda de dos entidades: nodos de decisión y hojas [12].

Las Redes Neuronales Artificiales (ANNs) también llamadas Redes Neuronales, es una estructura computacional que sigue a las redes neuronales biológicas. La capacidad de aprendizaje es la principal característica de las RNA. El proceso de aprendizaje se logra ajustando los pesos de las capas internas de acuerdo con algunos algoritmos de aprendizaje aplicados [13]. Support Vector Machine (SVM) es un enfoque popular de aprendizaje automático supervisado.

SVM se puede aplicar tanto para fines de clasificación como de regresión. SVM simplemente crea un hiperplano que puede dividir un conjunto de datos en dos clases diferentes [14]. Por otro lado, el El clasificador naive bayes requiere una pequeña cantidad de datos de entrenamiento para estimar los parámetros efectivos (medias y varianzas de las variables) que son necesarios para clasificar el conjunto de datos [15].

En este estudio, el algoritmo naive bayes no funcionó tan bien como otros cuatro algoritmos. El aprendizaje basado en instancias (Ibk) es un algoritmo que calcula las distancias de los vecinos más cercanos y luego clasifica el conjunto de datos en función de esas distancias. El número de vecinos más cercanos se puede cambiar si es necesario. Salomón

Mwanjele y et al. mostró que el clasificador Ibk genera

el error mínimo para predecir y el error cuadrático relativo de la raíz mínima. Lo usaron para predecir la precipitación por adelantado con mayor precisión en comparación con los otros dos clasificadores, a saber, la discretización y el algoritmo clasificador de regresión isotónica.

[dieciséis]. Como observamos en la situación actual, es obvio que si descubrimos el grupo objetivo de estudiantes que necesitan un asesoramiento adecuado en una etapa temprana, sería mejor para su vida académica y para su vida futura. Teniendo en cuenta todo eso, integramos la puntuación de factores con el algoritmo de aprendizaje automático para clasificar el conjunto de datos de los estudiantes para predecir el requisito de asesoramiento. Para desarrollar y mejorar la fuerza y la conciencia personal de los estudiantes, debemos guiarlos adecuadamente lo antes posible. En este trabajo trabajamos para predecir la orientación de los estudiantes para orientar su vida académica y para lograrlo realizamos los siguientes 3 pasos:

- Seleccionar funciones de acuerdo con nuestro objetivo y luego recopilar datos de estudiantes de pregrado
- Establezca una puntuación en cada función para obtener un nivel de clase más preciso
- Selección del modelo y luego evaluar el modelo con 10 veces técnicas de validación cruzada y división de datos del 66 %

II. MATERIALES Y MÉTODOS

En esta sección vamos a explorar nuestra propuesta enfoque para predecir un estudiante que necesita asesoramiento. Aquí nosotros se discutió sobre el procedimiento de recopilación de datos, el análisis de datos para la puntuación de factores, así como un diagrama de flujo para presentar el procedimiento de trabajo completo. También sería de apoyo para cualquier extensión adicional de este trabajo, así como para implementar cualquier aplicación.

A. Recopilación de datos y puntuación

Los datos de calidad, la cantidad de datos y la clasificación de datos son partes importantes para obtener una mayor precisión en cualquier enfoque de aprendizaje automático. Para este estudio, recopilamos 498 respuestas de los estudiantes mediante el formulario de Google. Fue un enfoque basado en una encuesta que trazó la condición de los estudiantes según el área de asesoramiento y todos los datos pertenecían a estudiantes de nivel universitario. Los datos se recopilaron del departamento de informática e ingeniería de la Universidad Internacional Daffodil, Dhaka, Bangladesh [17]. Consideramos seis atributos o factores para predecir el asesoramiento para un estudiante específico, descritos en la Tabla I. En el caso del procesamiento de datos, limpiamos algunos errores como nulo, desequilibrio de datos, etc. Luego establecimos un puntaje contra cada factor para establecer un nivel de clase/decisión según el asesoramiento necesario. Evaluó esta clase/decisión basada en la puntuación comparándola con la clase dada del estudiante y la clase dada por el consejero del estudiante y finalmente consideró un nuevo conjunto de datos con un valor de clase como un valor de clase final para predecir la clase.

Después de la selección de características, organizamos la Tabla II, que se preparó para representar los atributos con puntaje, a saber, SGPA, CGPA, SGPA vs. CGPA, Amistad, Área favorita, Debilidad y Tipo de dos clases (SÍ y NO), que se consideraron para hacer el conjunto de datos inicial. como describimos en la Tabla I y ese fue el segundo paso de este trabajo en papel porque el primer paso fue seleccionar el atributo y también encontramos la característica más efectiva usando el método de clasificación de ganancia de información.



Tabla- I: Descripción del Atributo

No.	Atributo	Descripción
1.	SGPA	Promedio de calificaciones del semestre
2.	CGPA	Promedio acumulativo de puntos de calificación
3.	Amistad	Pregunta de una sola línea "Sí o No"
4.	Favorito Área	Una pregunta de una sola línea con algún tema/área relacionada con el curso como- <ul style="list-style-type: none"> • Programación • Desarrollo web • Investigación • Juegos • Otro
5.	Debilidad	Una pregunta de una sola línea con alguna área- <ul style="list-style-type: none"> • Problema de comunicación • Arrogancia • No puntual • Problema de lenguaje • Irresponsable
6.	SGPA frente a CGPA	Observar la mejora actual según el rendimiento académico

Aquí, establecimos una puntuación alta frente a CGPA y SGPA porque nuestro objetivo era centrarnos en el asesoramiento académico más que en otros. Para calificar cada atributo, también analizamos los comentarios de los maestros como consejero estudiantil. Recopilamos datos del consejero de los estudiantes donde dieron una decisión sobre el nivel de clase de acuerdo con nuestros datos recopilados de estudiantes universitarios. Este enfoque de puntuación factorial se utilizó para proponer nuestro modelo para predecir el asesoramiento de los estudiantes.

Aquí, establecemos la clase de predicción (Sí o No) de la siguiente manera:

• Si la suma de todos los valores de los atributos es mayor que cero, entonces no hay necesidad de brindar asesoramiento a ese estudiante específico (sin clase).

• Si el valor es menor o igual a cero entonces el estudiante necesita consejería (Clase YES).

Ahora, podemos definir la definición formal del enfoque de predicción de consejería estudiantil de la siguiente manera:

Dado, m = número total de muestras de entrenamiento (X_1, X_2, \dots, X_m), $n=6$, que es el número total de características aquí que consideramos y es el estado de asesoramiento. Mostramos esta función en la ecuación

$$f(z_i) = \begin{cases} \text{Si, } \leq 0; \text{ sí} \\ > 0; \text{ No} \end{cases} \quad (1)$$

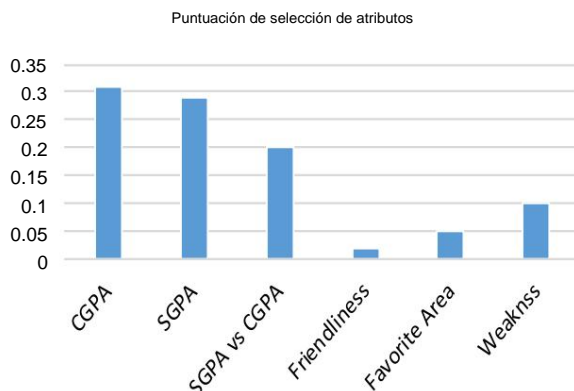


Fig. 1. Selección de atributos con puntaje

Cuando se aplicó el procedimiento de selección de atributos, obtuvimos CGPA como el atributo fuerte cuya contribución fue mayor que otros 5 atributos para predecir el estado de orientación de los estudiantes, ya que nos enfocamos en la orientación académica. Fig. 1 para mostrar los resultados de la selección de atributos utilizando la ganancia de información por método de clasificación. Este es un método para seleccionar el mejor atributo basado en el valor de ganancia de información y de esta manera también pudimos saber cuál proporciona más efectos en el nivel de clase.

Tabla- II: Puntuación de Factores Según Atributo

SGPA	si es menor que 2.5 entonces - 2	sí
	si 2.5 a 3.0 entonces -1	sí
	si 3.0 a 3.5 entonces 0	sí
	si 3.5 a 4 entonces 1	No
CGPA	si es menor que 2.5 entonces - 2	sí
	si 2.5 a 3.0 entonces -1	sí
	si 3.0 a 3.5 entonces 0	sí
	si 3.5 a 4 entonces 1	No
SGPA frente a CGPA	SGPA > CGPA = 0	No
	SGPA < CGPA = -1	sí
Amabilidad	Sí=0	sí
	No=1	No
Área favorita	Relacionado = 1	No
	Otro = 0	sí
	Nada = -1	sí
Debilidad	Sí = -1	sí
	No = 0	No
Tipo de clase	Valor total <=0	sí
	Valor total > 0	No

B. Análisis de algoritmos

En esta subsección, demostramos los resultados donde aplicamos cinco algoritmos populares y, al mismo tiempo, todas sus comparaciones de rendimiento de acuerdo con nuestro conjunto de datos. En la parte de evaluación del modelo, seguimos dos caminos para encontrar el modelo óptimo. Por otro lado, consideramos los comentarios de 5 maestros sobre el nivel de la clase como consejero y elegimos el valor promedio para establecer el puntaje para encontrar el nivel de la clase para hacer nuestro conjunto de datos. Aquí, cada maestro dio su decisión en contra de los requisitos de consejería de cada estudiante en base a 6 atributos. Se enfocaron en el asesoramiento académico ya que establecimos la primera prioridad en SGPA y CGPA para clasificar a los estudiantes. Después de eso obtuvimos el mejor modelo (algoritmo) para predecir la necesidad de consejería. Mientras tanto, aquí se utilizó la validación cruzada de 10 pliegues y el método de división del conjunto de datos del 66% para elegir el mejor modelo entre los 5 algoritmos de clasificación. La Tabla III va a presentar los resultados de dos observaciones entre 498 instancias. Los 5 maestros que recibieron comentarios se agregan aquí.

Algoritmo de aprendizaje automático y puntuación de factores para predecir la consejería estudiantil

Aunque no hay mucha diferencia entre ellos. Nosotros considerado su producción media. Todos estos miembros de la facultad fueron elegidos de Daffodil International University sobre la base de que asesoran a sus estudiantes durante al menos un año utilizando una plataforma particular especialmente diseñada para todas las facultades de Daffodil International University, para que puedan monitorear a todos los estudiantes y guiarlos para dar forma. sus futuros

Por lo tanto, podemos considerar su salida como estándar.

Tabla- III: Calificación de Atributos vs. Promedio del Docente Reacción

Considerado Parámetro (Estudiante Clase de consejería)	Puntuación de factores	Consejero estudiantil (SC = (T1 + T2 + T3 + T4 + T5) / 5)
SÍ	294	325
NO	204	173

Aquí, la Fig. 2 preparada para ilustrar el flujo de trabajo de nuestro modelo propuesto. Hay dos hechos observables, a saber, el consejero del estudiante y el método de puntuación que se consideró para proponer nuestro modelo final.

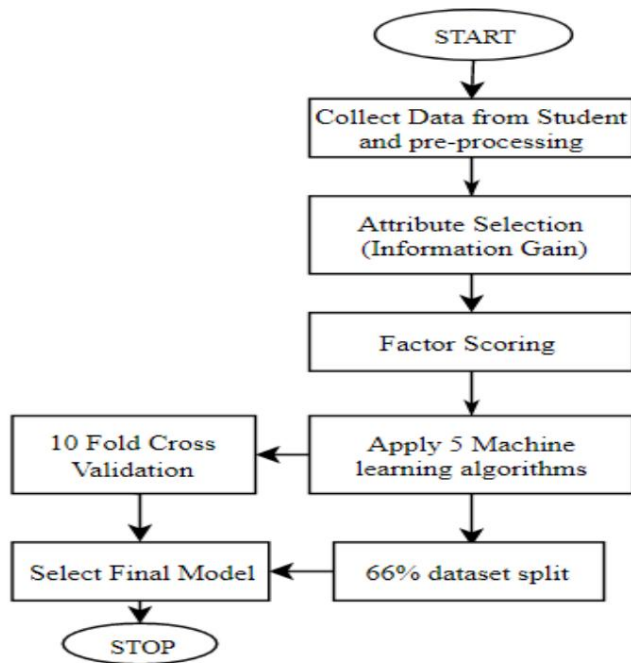


Fig. 2. Procedimiento de selección del modelo

Comparamos 5 algoritmos y encontramos un algoritmo que es lbk - "Conocimiento/aprendizaje basado en instancias" es más adecuado para nuestro enfoque. Aquí, también se describe el procedimiento para implementar este modelo en el futuro. En el caso de la implementación, el método de puntuación de factores se introducirá cuando llegue una nueva instancia al sistema después de compararlo con el valor umbral, entonces el sistema podría predecir el resultado final para el asesoramiento estudiantil como nuestra propuesta. En este documento, hemos elegido solo los cinco algoritmos de clasificación de aprendizaje automático más populares donde todos estos algoritmos proporcionaron resultados casi similares para nuestro enfoque propuesto.

Como mencionamos, lbk sigue al vecino más cercano se usa para algoritmo donde la función de distancia decide qué $D(X_i, X_j)$ vecinos (n) están más cerca de una instancia de entrada

de acuerdo con el cálculo de la distancia ponderada y también tienen un efecto intenso en un sistema de aprendizaje basado en instancias. Ecuación (2) para calcular la distancia de dos puntos.

$$D(X_i, X_j) = \sqrt{\sum_{r=0}^n (a_r(X_i) - a_r(X_j))^2} \quad (2)$$

Entonces, $D(X_i, X_j)$ = Distancia entre Xi y Xj.

Y luego considerada una función ponderada es la siguiente,

$$f(X_i) \leftarrow \frac{\sum_{j=1}^n w_j f(X_j)}{\sum_{j=1}^n w_j} \quad (3)$$

Donde, $w_j = \frac{1}{D(X_i, X_j)}$

C. Resultados

En esta sección discutiremos sobre el resultado. En la sección anterior encontramos que lbk es mejor para nuestro enfoque. lbk proporciona una precisión del 95,38 % según el método de validación cruzada de 10 veces. Las dos métricas básicas de desempeño, sensibilidad y especificidad, nos ayudaron a comprender la efectividad de nuestro método propuesto. Sensibilidad (Sn) y Especificidad (Sp), y precisión total definida de la siguiente manera:

Sensibilidad (Sn) = TP / (TP+FN)

Especificidad (Sp) = TN / (TN+FP)

Entonces, Precisión Total = (TP+TN) / N

Aquí, N es el número total de instancias. TP, FP, TN y FN son el número de regulaciones Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo respectivamente [18].

Verdadero Positivo (TP) es el caso donde el evento verdadero fue predicho correctamente. Verdadero Negativo (TN), donde el evento falso fue predicho correctamente. Falso positivo (FP) es el caso en el que se predijo incorrectamente un evento verdadero y, finalmente, Falso negativo (FN) en el que se produjo la predicción de un evento falso [19].

Para detectar un algoritmo más efectivo, utilizamos dos formas de analizar el conjunto de datos que se describe en la Tabla IV. Aquí presentamos la precisión de cinco algoritmos basados en una validación cruzada de 10 veces y una técnica de división de conjuntos de datos del 66 %. Es obvio que lbk generó el mejor resultado para predecir el asesoramiento para nuestro enfoque de puntuación de factores.

Tabla- IV: Comparación de Precisión

clasificador Algoritmo	cruz de 10 pliegues Validación (% de precisión)	División de datos del 66 % (Exactitud %)
lbk	95.38	94.08
bayesiana ingenua	85.14	94.08
multicapa	94.97	93.49
ESTAMOS	95.38	93.49
Bosque aleatorio	95.18	93.49

Validación cruzada de 10 pliegues, que es una forma en la que el conjunto de datos se divide en 10 pliegues y cada vez se prueba un total de 9 pliegues por 1 pliegue restante. Validación cruzada utilizada en tres casos: medición del rendimiento, selección del modelo y ajuste del parámetro del modelo. Mientras tanto, las divisiones del conjunto de datos del 66% son otra forma de análisis de datos donde el conjunto de datos total se divide en 2 partes, una parte contiene el 66% de los datos para el entrenamiento y el 33% restante para la prueba del modelo. Aunque las divisiones del conjunto de datos del 66% generaron casi los mismos resultados que la validación cruzada de 10 veces para los 5 algoritmos seleccionados, nos ayudó a diferenciar entre lbk y SMO, ya que ambos algoritmos generaron el mismo resultado según la validación cruzada de 10 veces. Después de usar la técnica de división de datos del 66 %, encontramos que lbk funciona ligeramente mejor que SMO.

También se creó la Tabla V para presentar 5 algoritmos seleccionados por valor de TP, TN, FP, FN para despejar el mejor modelo.

Tabla- V: Matriz de confusión de cinco algoritmos según dos categorías

clasificador	Cruz de 10 pliegues Validación		66% División de datos	
lbk	TP=293	FN=1	TP=106 FN=0	
	FP=22	TN=182	FP=10	TN=53
Bayesiano ingenuo	TP=242	FN=52	TP=106 FN=0	
	FP=22	TN=182	FP=10	TN=53
multicapa	TP=291	FN=3	TP=105	FN=1
	FP=22	TN=182	FP=10	TN=53
ESTAMOS	TP=293	FN=1	TP=105	FN=1
	FP=22	TN=182	FP=10	TN=53
Aleatorio bosque	TP=292	FN=2	TP=105	FN=1
	FP=22	TN=182	FP=10	TN=53

D. Discusión

En este documento, utilizamos nuestro propio conjunto de datos creado y no encontramos ningún estudio similar. Aunque muchos investigadores predicen el desempeño de los estudiantes en función de las actividades o características académicas, el asesoramiento fue raro. Algunos estudios se basaron en el asesoramiento profesional, pero nuestro objetivo fue proporcionar la necesidad de asesoramiento académico en función de 6 atributos seleccionados. En la sección anterior comparamos 5 algoritmos clasificadores para darnos cuenta de cuál es el modelo adecuado para predecir la necesidad de asesoramiento. En este documento, analizamos nuestro conjunto de datos de acuerdo con diferentes algoritmos en los que el asesoramiento de los estudiantes fue el punto central.

Tabla- VI: Desempeño de cinco clasificadores usando cinco métricas de evaluación

lbk	0.89	0.1	0.93	0.99	0.95
Ingenuo Bayés	0.89	0.1	0.91	0.82	0.86
multicapa	0.89	0.1	0.92	0.98	0.94
ESTAMOS	0.89	0.1	0.93	0.99	0.95
Aleatorio bosque	0.89	0.1	0.92	0.99	0.95

Hoy en día, los estudiantes están más ocupados con su vida social/medios de comunicación y otras actividades en lugar de estudiar como encontramos en muchos trabajos de investigación. Además, la tasa de deserción de los estudiantes ha ido aumentando día a día. Para ello, nuestro estudio fue conocer atributos que se relacionan con la orientación académica de los estudiantes. Sin embargo, lbk logró mejores resultados en nuestro estudio propuesto, también describimos todos esos 5 algoritmos clasificadores que son lbk, Naive Bayes, Multilayer, SMO y Random Forest con 5 métricas que se muestran en la Tabla VI.

Además, cinco métricas también funcionaron como parámetros de comparación en nuestro estudio presentado.

ROC (Características operativas del receptor) es un evaluador de rendimiento muy conocido. El área de la República de China también se ilustró en la Fig. 3. En esta figura, presentamos 5 áreas ROC de acuerdo con 5 algoritmos como mencionamos los datos en la Tabla 6. El área Roc proporcionó la proporción de Tasa de verdaderos positivos (TPR) que en realidad sigue la ecuación de Sensibilidad (S_n) y Falso positivo Tasa (FPR). Donde, $FPR = 1 - \text{Especificidad} (Sp)$. ROC es una forma más fácil de visualizar el rendimiento de los algoritmos de trabajo.

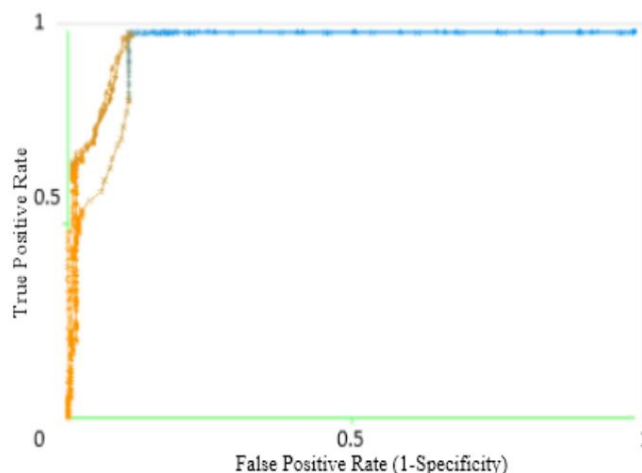


Fig. 3. Comparación del área ROC de cinco algoritmos

Finalmente, este estudio en papel nos ayuda a conocer la vida de estudio de los estudiantes y cuándo es útil brindarles asesoramiento con una mejor orientación y motivación para una vida exitosa. Mientras tanto, los estudiantes serán más conscientes de su vida y de sus padres si también compartimos esta información con ellos.

tercero CONCLUSIÓN

La frustración de los estudiantes en la vida universitaria no puede eliminarse por completo de todo el mundo, pero el nivel puede minimizarse si es posible predecirlo en la etapa inicial. En este documento, uno de los enfoques más cruciales fue identificar a los estudiantes procrastinados en el aula que necesitan un asesoramiento adecuado para una mejor vida estudiantil. Aunque, los atributos que se consideran aquí posiblemente no sean los únicos atributos para realizar este trabajo. Algunos de los otros atributos y pasos de procesamiento previo se pueden revisar y mejorar para lograr una mayor precisión. El resultado del consejero de los estudiantes y la opinión de los estudiantes, ambos se consideran para generar un conjunto de datos para este modelo. Se justifica la conclusión de que el algoritmo lbk es el predictor más fuerte entre los 5 algoritmos que se consideraron para este experimento.

Algoritmo de aprendizaje automático y puntuación de factores para predecir la consejería estudiantil

Por el contrario, encontrar la categoría adecuada de asesoramiento para un estudiante específico podría considerarse un trabajo futuro desafiante para ampliar nuestro trabajo de investigación.

Además, implementar este método y aplicarlo en la vida real es también nuestro próximo propósito de estudio. Si bien, se necesitarán más funciones e instancias para una creación de modelos más precisa.

RECONOCIMIENTO

Los autores desean agradecer al Departamento de Informática e Ingeniería de la Universidad Internacional Daffodil, Dhaka, Bangladesh, por alentarlos a completar este trabajo de investigación.

REFERENCIAS

1. El Daily Estrella. Disponible en línea: <http://www.thedailystar.net/news-detail-232009>, consultado el 12 de octubre de 2018.
2. Thaleia D., "Estudiante del Proyecto de Investigación de la Universidad Saludable". Encuesta de Salud y Estilos de Vida, junio 2015.
3. Nikita G., Ishani Z., Aishwarya N. y Deepali N., "Orientación profesional mediante minería de datos". Revista internacional de investigación innovadora en ingeniería informática y de comunicaciones, vol. 5, número 4, abril de 2017.
4. Pooja T., Anil M. y Manisha, "Análisis y predicción del rendimiento en la minería de datos educativos: un cuaderno de viaje de investigación". Revista Internacional de Aplicaciones Informáticas (0975 - 8887) Volumen 110 - No. 15 de enero de 2015.
5. Modo, y FN, "Servicios de consejería para el excelente desempeño académico de adolescentes en escuelas secundarias de la UYO". Akwa-Ibom. Actas de la conferencia de la Asociación de Consejería de Nigeria (CASSON), 2008.
6. Saiful I. y Nusrat J., "Digitalización y sistema educativo: una encuesta". (IJCSIS) Revista internacional de informática y seguridad de la información, vol. 16, N° 01, enero 2018.
7. Shruthi P. y Chaitra BP, "Predicción del rendimiento de los estudiantes en el sector educativo mediante la minería de datos". Revista internacional de investigación avanzada en informática e ingeniería de software. Volumen 6, número 3, marzo de 2016.
8. Emmanuel Awuni Kolog, Erkki Sutinen y Marjatta Vanhalakka Ruoho, "Implementación de asesoramiento electrónico: historias de vida de estudiantes y tecnologías de asesoramiento en perspectiva". Revista Internacional de Educación y Desarrollo utilizando Tecnologías de la Información y la Comunicación, (IJEDICT), 2014, vol. 10, Número 3, págs. 32-48.
9. Zamani, ZA. 2009, "Tecnología Informática y Asesoramiento. Facultad de Psicología y Desarrollo Humano". Facultad de Ciencias Sociales y Humanidades, IEEE.
10. N. Banu Priya, e I. Juvanna, "UNA APLICACIÓN DE ANDROID PARA LA CONSEJERÍA UNIVERSITARIA EN LÍNEA". Revista internacional de ciencias de la computación y computación móvil, edición Vol.3. 2, febrero de 2014, págs. 261-266.
11. David Fumo. Tipos de algoritmos de aprendizaje automático que debe conocer. Disponible en línea: <https://towardsdatascience.com/types-of-machine-learning-algorithms-usted-debe-saber-953a08248861>, consultado el 13 de octubre de 2018.
12. Jason Brownlee. Comprender los algoritmos de aprendizaje automático. Disponible en línea: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>, consultado el 16 de noviembre de 2018.
13. Sandeep K. y Medha S., "Convergencia de inteligencia artificial, inteligencia emocional, redes neuronales y computación evolutiva". Revista internacional de investigación avanzada en informática e ingeniería de software, volumen 2, número 3, marzo de 2012
ISSN: 2277128X.
14. Noel Bambrick. Máquinas de vectores de soporte: una explicación simple. Disponible en línea: <https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html>, consultado el 10 de octubre de 2018.
15. Harry Zhang, "La Optimalidad de Naive Bayes". FLAIRS 2004 Disponible en línea: conferencia.
<http://www.cs.unb.ca/profs/hzhang/publications/FLAIRS04ZhangH.pdf>.
16. Solomon Mwanjele Mwagha, Masinde Muthoni y Peter Ochieng, Comparación del vecino más cercano (1bk), Regresión por discretización

y algoritmos de clasificación de regresión isotónica para la predicción de clases de precipitación". Revista Internacional de Aplicaciones Informáticas 96(21): 44-48, junio de 2014.

17. Conjunto de datos del estudiante. Disponible en línea: <https://docs.google.com/spreadsheets/d/1Y26ywnUBN6J73AtltzSzPSSjtdgROfKz2ZSH0S7EwiA/edit?usp=sharing>, consultado el 21 de julio de 2018.

18. Colaboradores de Wikipedia. Sensibilidad y especificidad. Wikipedia, la enciclopedia libre. Wikipedia, la enciclopedia libre, consultada el 14 de marzo de 2018.

19. Jason Brownlee. ¿Qué es una matriz de confusión en el aprendizaje automático? Disponible en línea: <https://machinelearningmastery.com/confusion-matrix-machine-learning/>, consultado el 25 de junio de 2018.

PERFIL DEL AUTOR



Nusrat Jahan: Profesor de la Universidad Internacional Daffodil.

Completó su Graduación en CSE de la Universidad de Ciencia y Tecnología de Patuakhali, Bangladesh y una Maestría en Tecnología de la Información del Instituto de Tecnología de la Información de la Universidad de Dhaka. Tiene 4,5 años de experiencia docente.

Sus intereses de investigación son el aprendizaje automático, el análisis de datos y el procesamiento de imágenes. Tiene 4 publicaciones.



Saiful Islam: profesor titular en la Universidad Internacional Daffodil.

Completó su Graduación en CSE de la Universidad Internacional Daffodil, Bangladesh y una Maestría en Tecnología de la Información de la Universidad de Dhaka, IIT. Tiene 4,5 años de experiencia docente. Sus intereses de investigación son el análisis de datos, el aprendizaje automático

e interacción hombre-computadora.



Rezwana Sultana: completó su graduación y posgrado de la Universidad de Dhaka en Ciencias de la Computación e Ingeniería. Comenzó su carrera como profesora en la Universidad Internacional Daffodil en el año 2014. Actualmente trabaja como programadora asistente en el Departamento de Tecnología de la Información y la Comunicación de Bangladesh.

Su interés de investigación es la minería de datos, el aprendizaje automático y los algoritmos evolutivos.